DOCUMENT RESUME

ED 221 835 CS 006 827

AUTHOR Schadler, Margaret; Juola, James F.
TITLE The Development of Reading for Comprehension: An Information Processing Analysis. Final Report.
INSTITUTION Kansas Univ., Lawrence. Dept. of Psychology.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
REPORT NO NIE-TR-82-15
PUB DATE Mar 82
GRANT NIE-G-78-0179
NOTE 81p.; Appendices removed due to copyright restrictions.

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *Computer Oriented Programs; Elementary Education; *Eye Movements; Higher Education; Lexicology; Memory; *Metacognition; *Reading Comprehension; *Reading Processes; *Reading Research; Reading Skills; Visual Perception; *Word Recognition

ABSTRACT

This report summarizes research performed at the Universtiy of Kansas that involved several topics related to reading and learning to read, including the development of automatic word recognition processes, reading for comprehension, and the development of new computer technologies designed to facilitate the reading process. The first section explains how the development of automatic word recognition as a function of reading skill was investigated in seven experiments using Stroop-like tasks. It describes a longitudinal study of the development of automatic word recognition and how this skill is used by adults. It comments on the semantics effects in the development of this skill and examines the dimensions of lexical coding in Chinese and English. The second section reports on studies that examined the comprehension of brief passages, especially the forward and backward search processes in sentence integration. The third section reports on studies that present text in a rapid, serial, visual presentation format. The section emphasises (1) visual search and the reading of rapid, serial presentations of letter strings, words, and text; (2) such reading with and without eye movements; and (3) the role of comprehension and memory in reading rapid, serial presentations of text. Summaries and conclusions are provided in the final section. (HOD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED221835

FINAL REPORT

Grant No. G78-0179

The Development of Reading for Comprehension
An Information Processing Analysis

Technical Report Number NIE-82-15

Margaret Schadler

James F. Juola

University of Kansas
Lawrence, Kansas 66045

March, 1982

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

CS 006 827

Table of Contents

		Page
Abstract		1
Introduction		3
I.	The Development of Automatic Word Recognition	5
	The development of automatic word recognition and reading skill	5
	A longitudinal study of the development of automatic word recognition	12
	Automatic word recognition by adults	20
	Semantic effects in the development of automatic word recognition	24
	Dimensions of lexical coding in Chinese and English	29
II.	Comprehension of Brief Passages	32
	Forward and backward search processes in sentence integration	32
III.	Comprehension and Rapid Serial Presentations	61
	Visual search and reading rapid, serial presentations of letter strings, words, and text	61
	Reading with and without eye movements: A reply to Just, Carpenter, and Woolley	64
	Comprehension and memory in reading rapid, serial presentations of text	65
IV.	Summary and Conclusions	68
References		70
Appendices (copies of published articles summarized in text)		

Abstract

This paper is a summary of research performed at the University of Kansas pursuant to a grant from the National Institute of Education (Grant No. G-78-0179, 1978 to 1981). The research involved several topics related to reading and learning to read, including (1) the development of automatic word recognition processes, (2) reading for comprehension, and (3) the development of new computer technologies designed to facilitate the reading process.

The development of automatic word recognition as a function of reading skill was investigated in seven experiments using Stroop-like tasks. Age and reading skill level of the participants in these studies ranged from nonreading children through skilled college students. Results of cross-sectional studies are consistent with the data from an extensive longitudinal study. Results of the first studies show that interference with color naming in the Stroop tasks begins to emerge early in the process of learning to read, increases, then subsequently decreases. Strings of identical letters delayed color naming for children just beginning to learn to read. The interference from words, presumably reflecting semantic processing, begins developing early but does not peak until approximately the fourth grade reading level. These different sequences of development of interference from different graphemic and semantic distractor conditions suggest that word recognition is the result of a number of component processes that develop gradually as children acquire skill in reading. A recognition variation of the picture-word interference task provide clear evidence that children access category information about words incidentally to picture naming. The results of this study also provide evidence that automatic word recognition skills are different for adults and children. Furthermore, adults do not evidence facilitation effects from congruent color labels in the Stroop task while children do. Adults' processing of incidental words in naming and recognition tasks seems much more context independent than does children's.

Studies of reading for comprehension in adults included a study of how sentences are integrated into a coherent internal representation. Integration of meaning across sentences was investigated using antecedent-anaphor relationships in four reading tasks. It appears to be the nature of the context rather than the specific task demands

that determines whether a forward or backward search process is engaged. In general, it appears that various tasks demands do not actually influence the nature of integration processes so much as they affect how clearly integration time can be inferred from the response time measured.

A series of studies was carried out using a computer-controlled display to compare reading comprehension when text is presented in a page format versus a rapid, serial visual presentation (RSVP) format. In RSVP, one or more words are presented at a time to a single location on the computer screen. The results showed that reading RSVP text at rates between 200 to 1200 words per minute produces a level of comprehension no worse than that obtained when the same texts are read in a normal page format for equivalent periods of time. Further modifications of RSVP formats to make them more like eye fixation inputs produced no gain and even some loss in comprehension levels. RSVP methods seem to improve comprehension levels of less-skilled readers, whereas highly-skilled readers seem to comprehend RSVP and normal text equally well. More sophisticated RSVP methods might lead to a general improvement in reading efficiency and have wide applications as computers become more available to the public.

A study comparing logographic (i.e., Chinese) writing systems with alphabetic systems (i.e., English) focused on lexical access and word memory differences in two languages. Chinese logographs were shown to produce a much greater amount of visual information in memory than English words. English words were stored as a more integrated code of visual, phonemic, and semantic information.

Introduction

Learning to read involves acquiring some new skills specific to reading and adapting other previously learned cognitive skills for use in comprehending meaning presented through the abstract visual form of the linguistic code. Most importantly, reading involves integrating the appropriate component skills, both old and new, into a coherent functional process. We hold that the components of the reading process are interrelated and that the interaction is dynamic, changing as skill level increases and as the complexity of the reading task changes. Reading itself, then, must be considered a dynamic process and must be investigated as such. Our approach to the investigation of reading is comparative and developmental; we compare children and/or adults at different reading skill levels in order to determine the course of skill acquisition and the changes in the reading process as skill develops. We investigate different components of reading using a variety of tasks and by placing different demands on the system.

Typically, the beginning reader has already acquired the linguistic and conceptual knowledge needed to understand oral language and to interpret visual experiences. The beginning reader must learn a new language code that is based primarily on the visual-spatial symbols of written language rather than the auditory--temporal elements of the speech code. Much of beginning reading instruction is directed toward mastering the visual form of the code; the emphasis is on letter and word recognition. The ultimate purpose of this phase of reading instruction is the rapid and essentially automatic identification of words, a skill assumed vital to comprehending written text.

Since word recognition is considered so vital a component of reading and reading instruction, our research program has also had a heavy emphasis on the development of this skill, and this work will be reviewed in some detail. The second major emphasis in any reading instruction program is that of comprehension, the understanding of the units of text larger than the single word. Such units include phrases, sentences, and larger units including complete texts. We have also begun to investigate various aspects of the comprehension process using a variety of research tasks and approaches as would be warranted with a process of such complexity.

Following our review of the development of word recognition skills, we will discuss our work on comprehension processes. Finally we will summarize and integrate our research and suggest its implications for the understanding of reading and its development. When we present research that has been publised elsewhere, it is briefly summarized in the body of the text, and the published version is included in the appendix. Research that has not yet been submitted for publication will be described more fully.

I. THE DEVELOPMENT OF AUTOMATIC WORD RECOGNITION

Rapid word recognition is an effective skill for readers in that it enables attention to be focused on accessing lexical and semantic information about words rather than on the mechanical aspects of the recognition process. LaBerge and Samuels (1974) have suggested that it might be unnecessary to divert attention during reading from the processing of meaning to the recognition of individual letters and words once the recognition process becomes automated. The combination of the emphasis on decoding skills in reading instruction and the publication of LaBerge and Samuels' model of automatic word recogniion made it imperative that the development of automatic word recognition be investigated. Several investigators (e.g. Ehri, 1976; Guttentag & Haith, 1978; 1979; Posnansky & Rayner, 1977; Rosinski, Golinkoff, & Kukish, 1975) have investigated some aspects of automatic word recognition. The work to be reported here represents a major effort to study the development of automatic word recognition both longitudinally and cross-sectionally.

The Development of Automatic Word Recognition and Reading Skill

Margaret Schadler and David M. Thissen

Memory and Cognition,
1981, 9, 132-141

In the Stroop task a color name, when printed in ink of a different color, interferes with naming the color of the ink (Stroop, 1935, see also Dyer, 1973; Jensen & Rohwer, 1966 for reviews). Similarly, in the picture naming version (Rosinski, Golinkoff, & Kukish, 1975), the name of another exemplar from the same category printed on a pictured object interferes with naming the picture. These extremely reliable interference effects are usually interpreted as evidence that the reader automatically reads the word despite the fact that doing so delays and increases the difficulty of executing the naming response (e.g., Posner & Snyder, 1975). These effects are considered evidence of automatic word recognition, and Stroop-like tasks are frequently used to study the development of this skill

(e.g., Rosinski, 1977; Rosinski et al., 1975; Ehri, 1976; Guttentag & Haith, 1978, 1979).

LaBerge and Samuels (1974), whose seminal paper directed attention to the concept of automatic processing in reading, considered automaticity to be characteristic of fluent reading and proposed that it developed slowly with extensive practice. However, young children who have had little formal reading instruction evidence interference on Stroop-like tasks. Interference emerges as early as the first (Guttentag & Haith, 1978; Ehri & Wilce, 1979; Stanovich, Cunningham & West, 1981; Posnansky & Rayner, 1977) or second grade (e.g., Ehri, 1976; Rosinski et al., 1975; Schiller, 1966) on both the picture and color naming versions. The word recognition process would seem to become automated much earlier than would be expected from LaBerge & Samuels' model (e.g., Rosinski et al., 1975; Ehri, 1976; Guttentag & Haith, 1979).

The research to date suggests that Strooplike tasks may be useful in identifying changes in the way in which words are recognized as a result of learning or development. The patterns emerging in the research are (1) rapid increase of interference followed by a decrease, and (2) different developmental functions for the component processes. These results, however, must be qualified by the fact that age or grade has been used as the developmental variable when reading skill is actually the crucial factor. Since a few investigators have shown differences in amount of interference between skilled and less skilled readers within the same grade (Guttentag & Haith, 1978; Pace & Golinkoff, 1976; Ehri & Wilce, 1979; Ehri, 1976), research using reading skill as the developmental variable is needed. If we are to argue that the developmental sequence of interference effects is curvilinear, nonreaders, who should not show these effects, should be tested. Also, a group of older children should be tested to provide a reference for peformance at the upper end of the reading scale. The first experiment was designed to investigate the development of some of the components of word recognition as a function of reading comprehension level. The second and third experiments replicated and further explored an unexpected source of interference in the first experiment.

Experiment I

Method

The subjects were 140 children grouped according to their comprehension scores on the Gilmore Oral Reading Test. Stimuli were 72 rear-projected colored slides of capital letters in red, blue, or green print on a black background.

When the slides were projected onto the screen, each letter was approximately 1 x 2 cm. The stimuli were (1) the words RED, BLUE, and GREEN in the congruent color, e.g., the word RED in red print, (2) the same words in incongruent colors, e.g., the word RED in blue or green print, (3) the nonwords UBD, LGNE, and NDRBE, and (4) strings of 3, 4, and 5 X's. The reading test consisted of a list of simple words and the Gilmore Oral Reading Test (Gilmore & Gilmore, 1968).

The design was 6 (comprehension level) x 4 (stimulus condition) x 3 (response color) x 2 (sex of subject) mixed factorial. The six comprehension levels were: (1) non-readers who showed no ability to read any of the 14 list words nor the Gilmore Oral Reading Test (mean age = 5.6), (2) beginning readers who were able to read one or more of the 14 list words but obtained a comprehension score of less than 10 on the Gilmore Oral Reading Test (mean age 6.1), (3) first grade level readers who obtained comprehension scores of 10-14 on the Gilmore Oral Reading Test, a grade equivalency range of 1.3-1.9 (mean age = 6.8), (4) second level readers who obtained comprehension scores of 15-20, a grade equivalency of 2.1-3.4 (mean age = 7.2), (5) fourth level readers who obtained comprehension scores of 21-27, a grade equivalency of 3.8-5.8 (mean age = 7.6) and (6) the pre-sixth grade children, all of whom attained reading comprehension scores above 27 (grade equivalency 6.2 or higher; mean age = 11.2). A group of 20 first, second, and third graders who obtained reading comprehension scores of 27 or higher (mean age = 7.8) were excluded from these statistical analyses. The median of the six response times for each stimulus condition was the dependent variable.

Results and Discussion

Response times to the stimulus conditions and reading comprehension levels were analyzed using orthogonal polynomial trend analysis (Bock, 1975). The trend analysis indicated that the response times to the four types of interfering stimuli reliably followed different curves as reading comprehension developed.

The response time functions for both the nonwords and the X'S conditions were essentially concave downward. They differ primarily in that the response time curve for the X's condition accelerates more quickly to an earlier peak than does which curve for which nonword condition, which peaks later than decelerates more rapidly. The initial upward trend of these curves is evidence that both the X's and the nonwords interfere with the color naming response for children learning to read. Interference from nonword strings has been found previously (e.g., Guttentag & Haith, 1978) and was expected. Strings of X's, however, are

frequently used as the control condition and are assumed not to interfere with color naming. This result is further explored in the second and third experiment.

In contrast to the convex trends across increasing reading comprehension for the X's and nonword conditions, mean response times in the congruent conditions exhibit an essentially linear decrease across the entire range. There are two factors potentially contributing to the reading skill-related decrement in the congruent condition. One is the typical age-related improvement in response times, which is usually held to be a nonspecific effect of maturation and experience. The second is a facilitative effect of the congruent word on naming the color. Since the curve for response times in the X's condition reflects the development of interference rather than neutrality, the relative contributions of these two factors cannot be determined for these data (but see Experiment II).

Finally, the trend for the incongruent color-words is concave downward, as were the curves for the X's and nonwords, with peak response times for the Stroop stimuli occurring between the second and fourth grade reading equivalent levels. The incongruent color-word trend has a different orientation from the other curves: the mean response times for the incongruent stimuli increase as reading comprehension increases from nonreaders to about the fourth grade equivalent then decreases, but the pre-sixth graders' mean response time is no faster than that of the nonreaders. In contrast, the response times in the other conditions all subsequently decrease after the initial increase, such that these response times for the pre-sixth grade group are substantially faster.

In summary, interference with color naming in all conditions but the congruent words began to emerge during the early phases of reading skill acquisition. Each stimulus condition, however, showed a different sequence of development. Understanding this relatively complex condition-by-skill interaction depends, in part, upon understanding the nature of the interference in the X's condition as it relates to reading skill. The reliability and generality of this unexpected source of interference is explored in the next two studies, then the development of Stroop interference and word recognition will be discussed.

Experiment II

Strings of X's are frequently used as the neutral or control condition, particularly when the discrete trial procedure is employed (e.g.,Dalrymple-Alford, 1972). The only evidence in the literature that letters might cause

interference in the Stroop task is Stanovich et al.'s (1981) recent finding that single letters interfere with color naming for first graders, and this effect was relative to strings of X's, their control condition. The X's have been regarded as a means of presenting approximately the same amount of color in a neutral context as for a potentially interfering stimulus. A major purpose of Experiment II was to determine if strings of X's would interfere with color naming relative to rectangular patches, the other common control in Stroop tasks. The second purpose of this experiment was to investigate the development of facilitation from the presence of the congruent word.

Method

The subjects were 71 children classified as reading at the second, fourth, or sixth grade equivalent level. The equipment and procedure used in the first experiment were also used in this one. The slides were prepared in the same manner except that lower-case letters were used. In addition red, green, and blue rectangles were included. The test consisted of 9 practice trials and 90 test trials.

The design was a 3(comprehension level) by 5(stimulus condition) by 3(response color) mixed factorial. The reading groups were the grade equivalents of second, fourth, and sixth as defined in the previous experiment. Stimulus conditions were incongruent color-words, congruent color-words, colored X's, nonwords, and rectangles.

Results

As in the previous experiment, the response time data were analyzed using the orthogonal polynomial trend analysis. For the planned comparisons, each of the other four stimulus conditions were set against the colored rectangles condition. The trend analysis of response times across the three levels of reading comprehension indicated that the five types of stimuli have similar curves and that only the constant components were significantly different. The difference between levels for each of the letter stimulus conditions and the rectangle were all reliable.

Discussion

There is both a facilitative effect on color naming from the congruent words and interference from the X's. Although we cannot assume that the linear relationships found among the stimulus conditions in this experiment will be maintained as age and/or reading skill increases, there is no theoretical basis for predicting that the facilitative effect found for these reading skill groups should disappear

in adulthood. Rather the relative advantage for the congruent pairs over the rectangles should be maintained for college students, whose word processing should be quite efficient. Since there is some evidence of greater deceleration in the response time curves for the X's condition than for the rectangles condition in this experiment and in the congruent condition in both experiments, it is possible that X's may not reliably interfere in color naming by older and more highly skilled readers. However, the effect can and should be tested empirically, particularly considering the amount of Stroop research conducted with college students.

Experiment III

The third experiment was designed to explore the development and the generality of the interference found in the X's condition. To do so, we included colored rectangles and pronounceable and unpronounceable nonwords. Contrary to the tradition of the Stroop task, all words were excluded in order to eliminate any possible effect their presence on some trials might have on the other stimuli. If beginning readers can ignore letters, they should be most likely to do so when there are no words in the entire test array.

Method

Subjects were 32 kindergarten children plus 14 girls and boys tested within a month of entering the first grade. The equipment and procedures were similar to those used in the first two experiments. Stimulus slides were prepared using the same photographic procedure as that used previously. The symbolic letters were x, o, and z. The common letters were e, s, and t, and the uncommon letters were k, p, and v. All letters were presented in strings of four. The pronounceable nonwords were mafe, stib, and hund, and the unpronounceable nonwords were sbti, mhua, and dnfe. The reading test consisted of three parts: a list of the letters used as stimuli, the list of words used in the previous experiments, and the Gilmore Oral Reading Test. These were given following the color naming and color vision tasks.

Design The design was a 3(comprehension level) by 6(stimulus condition) factorial. The reading comprehension levels used in this experiment were (1) ability to name most of the letters but none of the words, (2) reading time for the word list of greater than 20 seconds, and (3) a word list reading time of less than 20 seconds. There were 15 children in the letters--only group (mean age = 5.6 years), 22 slow word readers (mean age = 6.0 years), and 9 in the fast word reading group (mean age = 6.4 years). The use of

the letter and word reading lists enabled us to use the same measure to classify all groups.

Results

None of the main effects and interactions involving color were significant for the polynomial trend analysis. In the first planned comparison of the symbols against the common and uncommon letter strings there was greater interference from the symbols but only for the group who could name letters but not read words. Letters having both symbolic and alphabetic value may be more meaningful to young children and perhaps be recognized earlier than other letters. More interference was observed for the letters than for the nonwords, and response times for the letters tend to decrease across reading levels, whereas there is no overall linear change in response times for the nonwords as a function of reading level. Most importantly, the interference from letters both develops earlier and decreases prior to the comparable changes for nonwords. This difference is comparable to the results of the first experiment in which the interference for the X's both emerge and declined at lower levels of reading skill than did interference from the nonwords.

Discussion

Beginning readers do name colors more slowly when the interferent is a string of identical letters than when it is a nonword or a rectangle, and letters that have smbolic value interfer more with color naming by nonreaders. We have no idea of how much exposure to letters is necessary to produce interference in color naming, but it would seem to be much less than LaBerge and Samuels (1974) originally implied would be necessary for automaticity.

Why do individual letters produce interference for the child at this low level of reading skill? For some period in the beginning of reading instructions, individual letters may be the most salient entities for the child. Although some words may become highly familiar and thus automatically processed equally early in the course of learning to read, the letter is the common element. For the beginning reader the letter is the smallest element to be identified and, usually, the largest to be processed as a unit, thus, it is distinctive.

General Discussion

The results of these three experiments are remarkably consistent. When interference from mixed letter strings first emerges, there seems to be little difference among

words, pseudowords, and unpronounceable nonwords (see also Guttentag and Haith, 1978). This result suggests that the letter strings themselves are the source of the interference, the semantic properties of the words are not yet a significant factor.

By the second grade reading level, the incongruent words caused more interference than the nonwords. This difference as well as the difference in amount of interference between intra- and extracategory words noted in the picture-word task (e.g. Guttentag & Haith, 1978; Rosinski, 1977) indicates that words come to be processed to a greater depth and their semantic content accessed as reading skill increases. We suggest that the increasing portions of the Stroop interference curves are due to increasing familiarity with and organization of the components of word recognition. As such knowledge develops, the depth to which the interfering word can be efficiently processed increases. Deeper processing places more demand on the cognitive system, and the result is more interference in the color naming task.

A Longitudinal Study of the Development of Automatic Word Recognition

Margaret Schadler, David M. Thissen, and Donna Fisher

Most of the research on the development of skills that are acquired over a period of years is collected in cross-sectional paradigms. The assumption is made that data generated from sampling different children at designated skill levels are equivalent to those obtained from the same children tested at those same skill levels. Such an assumption is not always tenable; furthermore, cross-sectional methods allow only the determination of differences not changes. Recognition of the limitations of cross-sectional research necessitates validating longitudinal research on skill development. The present study was a longitudinal investigation of the development of automatic word recognition, using the Stroop task.

There have been only two other longitudinal investigations using Strooplike tasks. Both of these (Guttentag & Haith, 1980; Stanovich, et al, 1981) tested just first-grade children and limited the duration of the study to one year. The results of both of these studies are more relevant to the development of letter recognition and to early word recognition than to the broader range of

reading skill development that we encompassed. This longitudinal study had two major purposes: comparison of the results of cross-sectional and longitudinal studies of the development of interference in the Stroop task and exploration of the relationship between the development of reading skill and automatic word recognition.

Method

The potential sample consisted of the 120 kindergarten through third grade children who participated in Schadler and Thissen's first experiment. Of these, 103 participated in the second year, and 72 in the third. The materials and apparatus used were identical to those used by Schadler and Thissen (1981), with the exception that the C form of the Gilmore Oral Reading Test was given during the first and third years, and the D form was given in the second year. The procedure used in all three years of the study was identical to and is described by Schadler and Thissen.

Results and Discussion

Prior to formal analyses, we made a simple comparison of the cross-sectional and longitudinal data. The longitudinal sample was divided into two groups, those who could not read well enough to produce a measurable score on the Gilmore test (n=30) and those who could (n=73). All children who were retested at least once were included in one of these two groups. We then plotted the mean response times to each of the four stimulus conditions at each time of measurement for each group. These means are presented in Figure 1. For comparison purposes, the two groups are spaced 1.5 years apart, which is consistent with the mean difference in the ages of the two groups.

As can be seen from a comparison of Figure 1 and Figure 2, which depicts the cross-sectional data in Schadler and Thissen's first experiment, the patterns of response times times rapidly diverge from the other conditions in an essentially linear fashion. The initial curvature for the longitudinal representations of the response times in the X's and nonword conditions appears flatter than for the cross-sectional representation; this kind of flattening could easily result from the grouping of wider ranges of reading skills in the longitudinal sample during the second and third years of the study. There are slight disjunctions in the curves for the Stroop and congruent response times between the last year for the nonreading group and the first year of the reading group; these do not seem particularly important. The essential point to be noted here is that cross-sectional studies of the development of interference from in the Stroop paradigm appear to be

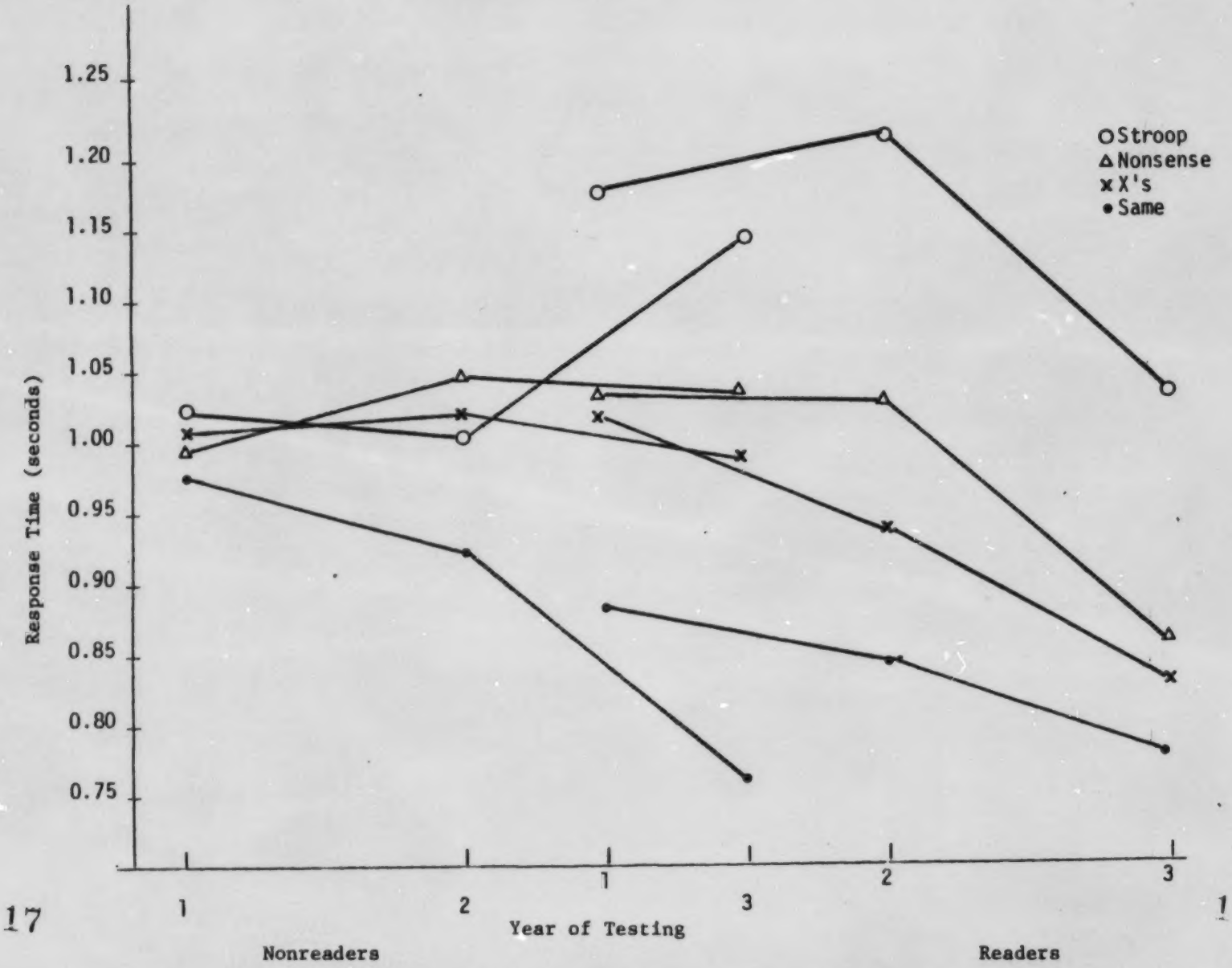

Figure 1. The development of interference on the Stroop task as a function of reading skill: Longitudinal data

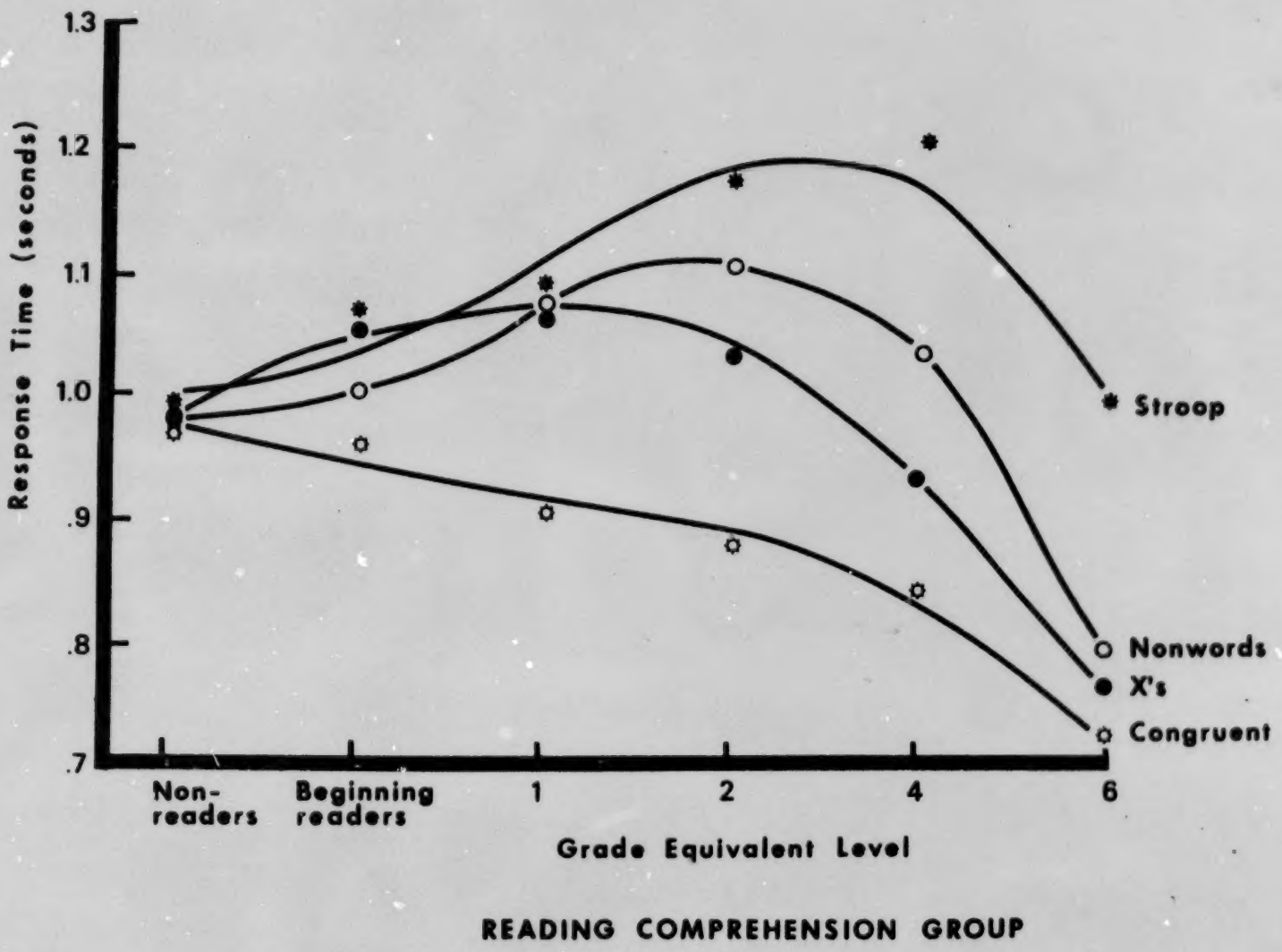

Figure 2. The development of interference on the Stroop task as a function of reading skill: Cross-sectional data

acceptable representation of changes that occur longitudinally.

The individual differences data from the three-year longitudinal study were analyzed with the techniques of path analysis, more recently called the analysis of linear structural relations by Joreskog (1974, 1978; Joreskog & Sorbom, 1977, 1978). The data have the form of a "panel study"; within each of the three waves of testing, six variables to be used here were derived from the data for each child. Two of these, the Gilmore Oral Reading Test Accuracy and comprehension scores, were taken to be "congeneric" (Joreskog, 1974) measures of reading ability. The other four variables obtained at each occasion were the average response times for the four conditions of the Stroop task: congruent, incongruent, nonwords, and X's.

For this analysis, the complete data for the three-year longitudinal study are summarized in an 18x18 covariance matrix, containing the variances and covariances among the six variables measured at each of the three occasions. See Table 1. Complete data to make up such a matrix were available for 55 of the 120 participants in the study. Only incomplete data were available for the remaining participants as they either dropped out after one or two years of testing; or they simply could not read during the year one testing and thus obtained no score at all on the Gilmore scales at that time. The participants with incomplete data do, however, provide information about the relationships among the variables for which they do have data; this information may be represented in smaller covariance matrices subsets of the complete covariance matrix.

We created a "path model" or set of linear structural equations describing the relationships among the variables at the different occasions (the model is illustrated in Figure 3) and obtained the maximum likelihood estimates of the parameters of that model using all of the data, complete and incomplete alike, and the computer program LISREL IV (Joreskog & Sorbom, 1978). The theory of maximum likelihood estimation also provides a goodness of fit test for the model, and this model was found to be a satisfactory (if complex) representation of the data.

The Gilmore scores did indeed provide congeneric measurement of reading ability; the accuracy score provided better measurement than did comprehension. Individual differences on reading ability were very consistent over the three years of the study.

Table I

Correlation Matrix for Longitudinal Data

(Decimals Omitted)

Year		1		2		3		1				2				3			
		A	C	A	C	A	C	S	N	X	UC	S	N	X	UC	S	N	X	UC
1	A	-																	
	C	7	-																
2	A	8	7	-															
	C	6	6	7	-														
3	A	6	3	8	6	-													
	C	6	5	7	7	7	-												
1	S	0	0	1	1	-0	-2	-											
	N	-1	-0	-1	-0	-1	-2	9	-										
	X	-3	-3	-3	-3	-3	-4	7	8	-									
	UC	-2	-2	-2	-1	-2	-4	7	8	9	-								
2	S	-4	-2	-3	-3	-2	-3	5	5	6	5	-							
	N	-6	-4	-5	-4	-4	-5	4	5	6	6	8	-						
	X	-6	-3	-5	-3	-4	-4	5	5	7	6	8	9	-					
	UC	-4	-2	-3	-3	-3	-4	5	5	6	6	4	8	9	-				
3	S	-5	-3	-4	-3	-3	-3	3	4	5	5	6	7	6	6	-			
	N	-6	-4	-6	-4	-4	-5	4	5	7	6	7	8	8	6	7	-		
	X	-5	-4	-6	-4	-4	-4	3	4	7	6	5	8	8	6	7	9	-	
	UC	-3	-2	-2	-1	-2	-2	4	4	6	6	8	6	6	5	7	8	8	-

Where A = Achievement
C = Comprehension
S = Stroop
N = Nonword
X = X's
UC = Congruent

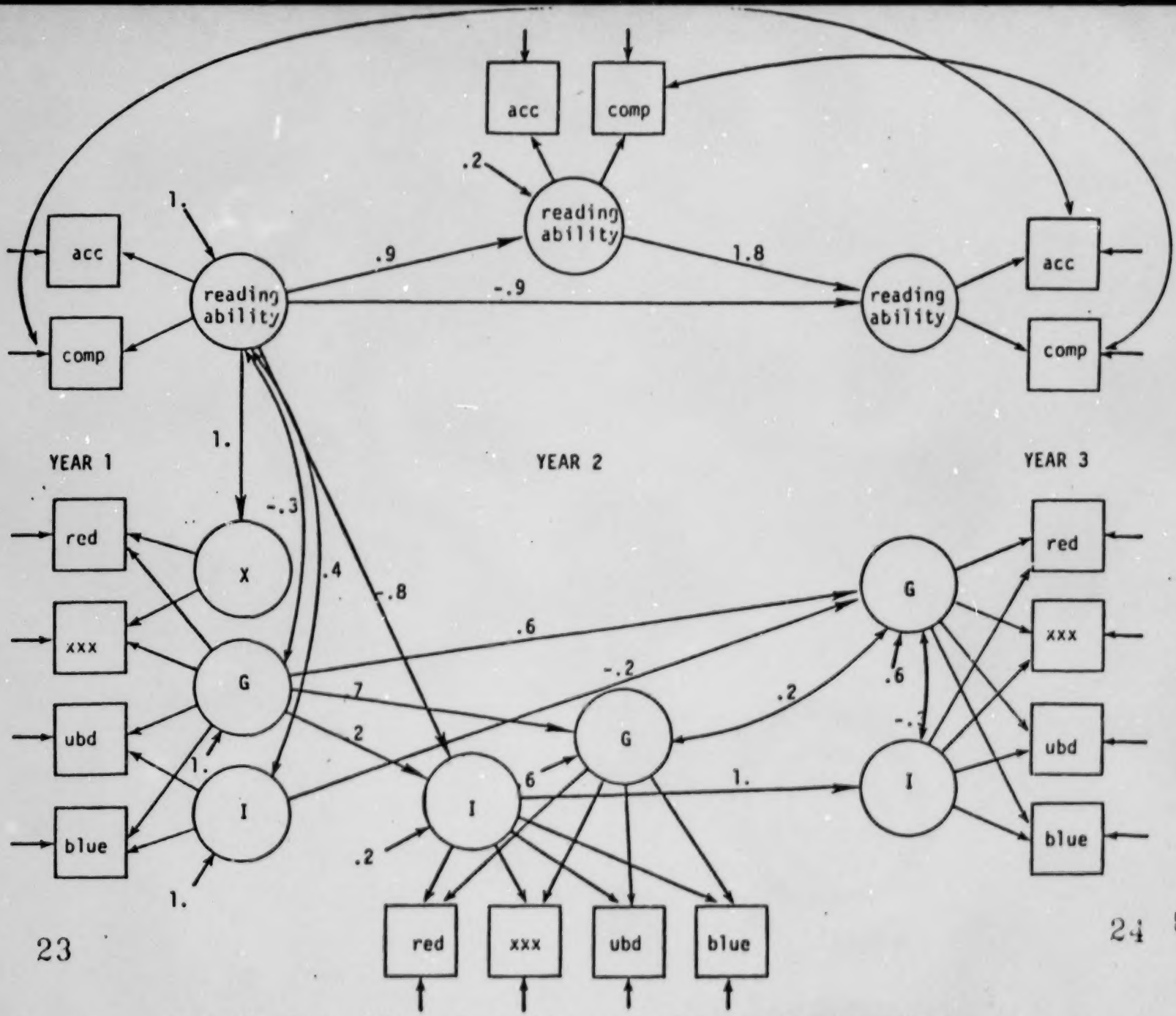

Figure 3. Path model of relationships among reading and Stroop variables

The response times for the Stroop task reflected three underlying dimensions of individual differences in the first year and two in each of the second and third years. The first dimension is general speed of response, indicating a consistency in individual response speed across stimulus conditions and across the three years of testing. The relationship between this factor and reading indicates some tendency for better readers to respond faster than less-skilled readers.

In the first year, there were individual differences in the extent to which X's and congruent color labels affected color naming, and this dimension was strongly related to reading ability. The better reader shows greater difference in their response times in these two conditions than the less-skilled reader. This rather complex relationship is consistent with our cross-sectional finding that interference in the X's condition emerged and peaked early in the process of learning to read. Poorer readers would show little difference. They don't automatically and differentially encode the X's and color names. Reading ability in the second and third year, however, is predictive of performance in the X's and nonword conditions, and the relationship with amount of interference is reversed. During these later years, the less-skilled readers are showing relatively more interference in these conditions than are the better readers. Said another way, interference from the X's and nonword conditions has become relatively minor for the more skilled readers but remains a problem for the less-skilled. It may be that the less-skilled reader has more difficulty in processing nonwords as such than the better reader. These effects are consistent with results by Perfetti and Roth (1982) and Fredrickson (1982) that show that, for older readers, response times for naming unfamiliar and pseudowords is more highly related to reading skill than is naming real or common words.

In sum, early reading ability (year 1) was highly predictive of later reading skill. The effect may be exaggerated in this study because the Gilmore test is a relatively simple instrument and neither the form of the test nor the type of material to be read change much across grade levels. By the same token, reading ability in the second and third years made no significant contribution to predicting performance in the Stroop task. After showing strong relationships early in the process of learning to read, reading skill and Stroop performance diverged, follwing their own internally consistent paths. Relationships between reading skill and the Stroop tasks obtained late in development were attributable to the earlier relationship between them and the consistency of each over time.

Automatic Word Recognition by Adults

Margaret Schadler

In their second experiment Schadler and Thissen (1981) found, using the Stroop task, that children named the color of strings of 3-5 X's reliably slower than they named the colors of rectangles that were the same length as the four-letter strings. The children were 71 second, third, and fourth graders who were reading at second to sixth grade equivalent or higher. In their third experiment Schadler and Thissen found that, at least for nonreading and beginning reading children, other identical letter strings also delayed color naming relative to the rectangles. Children reading at the first grade level named colors of identical and mixed letter strings reliably slower than they named the colors of rectangles. The present experiment was designed to determine if adults also showed an interference effect for letters in the Stroop task such that they would name colors of strings of X'S more slowly than they would name the colors presented in the rectangular shape.

Method

Subjects. The subjects were 15 college students who participated in the study in partial fulfillment of their introductory psychology course requirements. All subjects were native speakers of English and had normal or corrected to normal distance vision and normal color vision.

Materials and procedure. The materials and procedures were the same as used in Schadler and Thissen's second experiment. The same stimulus conditions were used; the distractors were congruent and incongruent color names, strings of X's, unpronounceable nonwords. Colored rectangles, the same length as the 4-letter strings and half their height, served as the basic comparison (control) condition. The stimulus colors were red, green, and blue. The task consisted of 9 practice trials and 90 test trials presented in 6 blocks of 15 trials each. Three exemplars of each stimulus condition appeared in each block of trials. The stimuli were presented as colored forms on a black ground, rear-projected by a 35mm slide projector. A digital timer was started when the slide appeared and stopped, using a voice operated relay, when the subject named the color. Each subject was tested individually. Prior to administering the Stroop task, the experimenter used Ishihara Plates 24, 25, 28, and 29 to test color vision.

Results and Discussion

The data were analyzed using analysis of variance with one repeated measure, stimulus condition. The effect was reliable; $F(4,11)=52.36$, $p<.001$. Orthogonal comparisons showed that the mean response time in each of the other conditions was reliably slower than the mean in the rectangles condition. The mean response times for each condition in msec are incongruent, 856; nonwords, 715; congruent, 699; X's, 684; rectangles, 665. As we had hypothesized, the mean response time for the X's condition The difference between mean responses in the X's and rectangles condition is small but reliable ($p = .026$). It is also consistent with the results of the two studies with children. This interference effect is evidence that even skilled readers are susceptible to the effects of the semantic context of the strings of X's. While experimenters may consider the X's to be simple place-holders or means of equating the amount of color or light in the control and experimental conditions, subjects do not. The X's have meaning, and readers process it. The utility of X's or any other shape for use as a control condition for this task should be evaluated carefully in terms of the purpose it must serve. When the control must serve to potentially determine facilitation as well as interference, selection becomes particularly important.

There is, however, a major difference between the results of this study and that of our work with children. The adults' mean response time in the congruent condition is also reliably slower than in the rectangles ($p = .006$); it is slightly but not reliably slower than in the X's condition. While the congruent labels reliably facilitated color naming relative to both the rectangles and the X's for children; they apparently delay color naming for the college students. Since this was an unexpected effect, we decided to test the strength of the effect in a second experiment.

Experiment II

Method

The subjects were 28 college students, all native English speakers with normal or corrected to normal vision and normal color vision. The materials and procedures were the same as used in the previous study with two exceptions. The rectangles were smaller; the ones used in this study were the same length as the 3-letter strings and half their height. We also added a set of pronounceable pseudowords, (lud, nege, drube). These pseudowords were constructed from the letters comprising the three color names with the restriction that the first letter could not be the first

letter of one of the color names. There were 10 practice and 108 actual trials. Exemplars of each stimulus condition appeared in each block of trials.

Results and Discussion

The median score for each person's response to each distractor type in each color was calculated, then these were analyzed in a 6(distractor type) x 3(color) analysis of variance. The means are presented in Table 1. The main effect of distractor type was reliable $F(5,135) = 121.31$, $p < .001$. Mean response time in the X's condition was slower than in the rectangles condition, but there was no reliable difference between the congruent and rectangles conditions. Thus the interference effect of the congruent words found in the previous experiment was not obtained here with the smaller rectangles and the introduction of the pseudoword condition. One or both of these stimulus changes may account for the differences in outcome or there may be additional factors that have not been identified. In any case, selection of control conditions should be given careful attention when age-related differences in automatic word recognition is being investigated. This concern is particularly important when the issue of facilitative effects is important. The issue of facilitation in the Stroop task, particulary from the congruent words, has been investigated. The results have indicated that the congruent words do facilitate responding but the control stimuli have been letters or geometric symbols (cf. Dyer, 1973).

There was a main effect of color, $F(2,54) = 14.27$, $p < .001$, and a color by condition interaction, $F(10,170) = 3.58$, $p < .001$. In the X's, rectangles, and Stroop conditions, mean response times to red and green were faster than responses to blue. In the pseudo- and nonword conditions, red was named fastest, blue intermediate, and green slowest. In the congruent condition, red was named faster than blue and green. Differences in response times to color are consistent across experiments only in that red is named as fast or faster than either of the other colors. The condition by color interaction is not always reliable nor is the relationship between color and distractor word condition consistent when it is reliable. Identifying the factors affecting this relationship is beyond the scope and purpose of this study. The importance of the color variable is largely as a control factor.

Table I

Mean Response Time on the Stroop Task

Experiment II

	Red	Blue	Green
Stroop	830	819	873
Congruent	588	626	627
Pseudoword	614	666	696
Nonword	616	635	657
X's	622	648	620
Rectangles	591	624	604

Semantic Effects in the Development of Automatic Word Recognition

Margaret Schadler

Our previous work on the development of automatic word recognition using the Stroop task has been largely focused on the grapho-phonemic and unspecified meaning components of word recognition. In the present research we shift our emphasis to the semantic content of words, specifically to categorical information contained in word meaning. Smith and Magee (1980) have shown that adults access category information for both distractor and target stimuli on the picture-word interference task. Lupker and Katz (1981) found similar results using a discrete trials procedure in which college students responded "yes" or "no" on each trial depending upon whether or not the object pictured was a dog. Names of other animals superimposed upon the picture of a dog delayed the positive response whereas names of nonanimals did not relative to control pictures that had no label. Labels did not reliably affect the speed of responding "No" to pictures of animals othe than dogs and to nonanimal pictures. There was a marginal delaying effect when the word "dog" was superimposed on the nondog pictures.

The Lupker and Katz task provides an opportunity to investigate the development of a semantic component of automatic word recognition. Rosinski (1977) and Guttentag and Haith (1978) reported more interference from names of objects in the same category upon picture-naming than from names of objects from different categories. However, the possibility exists that much, if not all, of the interference in these studies is due to activation of the name code. Interference in the picture recognition task is greatly reduced for adults (Lupker & Katz, 1981; Smith & Magee, 1980). The task also provides an opportunity to evaluate relative amounts of interference across age as a function of category distance among the target basic category, its superordinate, and the broad category of outline drawings of objects.

The subjects are second and fourth grade children and college students. The study was conducted to explore potential age-related differences in the type of category information accessed in a picture-word interference task in which the subject is asked to identify whether the object pictured is a dog or not.

Method

Subjects. The subjects were 15 second and 15 fourth grade children from Lawrence, Kansas. Each child was paid $2.00 to come to the University of Kansas campus where s/he was individually tested in a laboratory room in the Psychology Department. Fifteen college students also participated, receiving research credit in partial fulfillment of their introductory psychology course requirements. All subjects were native English speakers, had normal or corrected to normal near distance vision and normal color vision. Each college student and a parent of each child gave written consent prior to participating. The children gave oral consent.

Materials and Procedure. The apparatus is one we have used previously for investigations of automatic word recognition using Stroop-like interference tasks. A 35mm Kodak Carousel projector is used to display individual picture-word combinations on a 27.5 cm square rear-projection screen. A solid state timer is started with a photo-transistor when a stimulus slide is projected and terminated with the subject's oral response, using a voice-operated-key. Participants were instructed to name the object pictured as rapidly as possible without making errors. They were told to ignore the letters. A verbal ready signal was given prior to the onset of each picture. Verbal feedback and encouragement were given for correct responses; the subject said "yes" if the picture was that of a dog and "no" to any other object. If any errors were made, the participant was asked to correct it before the display was terminated. Between trials, the experimenter recorded the response time and any error made.

The stimulus slides were prepared using Kodalith 35mm film to photograph black-on-white outline drawings with the letters or symbols superimposed on the center of the drawings. The projected drawings were approximately 6 cm square with appropriate adjustments for the proportions of the animal or object depicted. When projected each letter or symbol was approximately .4 by .6 cm. The positive instances were outline drawings of 10 different kinds of dogs in various poses, e.g., standing, sitting, obliques, face on, in order to produce slides as different as possible. Half the negative instances were other animals (bear, mouse, cat, horse, fish) and half were nonanimals (foot, apple, leaf, shoe, table). The distractor items were the word "dog," the names of the other animals, the names of the nonanimals, and visual noise strings consisting of the symbols &(%$) and &?!$. The picture and distractor word were congruent only in the positive condition, never when the picture was an object other than a dog.

Participants were tested individually following a brief introduction, establishment of rapport, and explanation of the task and its purpose. Each person was given 8 practice and 80 test practice trials.

Results

The response time data were analyzed in a 3(grade) x 3(picture type) x 4(word type) mixed analysis of variance. Grade (second, fourth, college) was a between-subjects factor; picture type (dog, other animal, nonanimal) and word type (dog, animal name, nonanimal name, visual noise were within-subject factors. Each subject's median response time in each condition was the basic datum. The means are presented in Table 1. The main effect of grade is significant as expected; $F(2,42) = 87.03$, $p < .001$. Older subjects respond faster than younger ones. The main effect of picture type is significant, $F(2,84) = 17.76$, $p < .001$. Mean response times are longer when the pictures were animals other than dogs and the subject must respond negatively. Response times are essentially identical when the subjects responded positively to pictures of dogs and negatively to pictures of nonanimals. However, this analysis confounds stimulus type and response type; responses to positive stimuli are typically faster than to negative ones. Since there are also significant higher order interactions involving picture and distractor type, this main effect will not be discussed. The main effect of distractor type is marginally significant ($p = .059$) but the confounding of distractor type with word type and therefore response type renders it meaningless as a main effect.

The picture x word x grade main effect was reliable, $F(12,252) = 1.85$, $p = .04$, as was the picture x word interaction, $F(6,252) = 8.90$, $p < .001$. When the picture is a dog and the response is positive, all three grade levels perform similarly, responding faster when the distractor is the word dog, somewhat slower when it is visual noise, even more slowly when the word is the name of a nonanimal, and slowest when the word is the name of an animal other than a dog. These are the expected results and consistent with Lupker and Katz's (1981) results. When the picture is not a dog and the response is negative, we begin to see some age-related differences in response patterns. When the picture is an animal other than a dog and the distractor is a name of a nonanimal, second graders responded quite slowly. is a name of a nonanimal. The same pattern is present for the fourth graders but the size of the effect is negligible; the adults do not show it. There is little difference in this animal picture condition between the noise and dog distractors. All three groups respond to the animal pictures fastest when the distractors are the names of other

Table I

Mean Response Times on the Picture Recognition Task

	Grade 2	Grade 4	College
Dog			
Dog	107	089	067
Animal	123	104	073
Nonanimal	118	097	071
Visual Noise	111	095	068
Animal			
Dog	127	108	077
Animal	115	095	075
Nonanimal	138	110	075
Visual Noise	127	105	074
Nonanimal			
Dog	126	099	065
Animal	116	092	064
Nonanimal	117	100	067
Visual Noise	117	097	065

animals; however, the effect was relatively larger for the children than for the adults. Finally, in the nonanimal picture condition, the second grader's responses were slowest when the distractor was the word "dog"; this interference was not evident for either the adults or the fourth graders.

Discussion

The children and adults appear to perform the positive recognition phase of this task in much the same way. When the picture is that of a dog, all three groups respond similarly. Further these data are comparable to those obtained by Lupker and Katz (1981). These results are evidence that both children and adults access category and word information during the course of this task even though the semantic information about the words is irrelevant to the task and doing so delays the recognition response when the category of the word is not the same as for the picture. These interference effects are consistent with those found in Stroop-like naming tasks.

The most generally accepted explanations of Stroop-like tasks are limited channel capacity models of attention (see Dyer, 1973; Posner & Snyder, 1975). These models typically assume independent processing of pictures and words until the response stage, then the automatically processed item competes with the more slowly intentionally processed response. As a consequence of the competition, the response is delayed; the greater the similarity between the responses, the greater the interference.

The children's performance in the negative conditions is not consistent with the response-conflict automaticity model The children seem to process the picture and word components of the task in a highly interdependent fashion. That is, they seem to process the distractor word to the relevant category level before they will make a decision about the category of the picture; there is no other explanation for the greater interference from nonanimal labels than from animal labels. We suggest that the animal names and pictures are primed through their superordinate relationship to dog thus facilitating recognition of exemplars of these concepts. The animal names are congruent with animal pictures at the category level thus facilitating their recognition, relative to the visual noise category, but have little effect on the nonanimal pictures. This pattern is what occurs in the children's response pattern. Further, nonanimal names are not primed thus they have no facilitating effect for their own category and may even cause some delay in the other conditions because they must

be accessed and categorized before the response is made; such a delay is evident particularly for the second graders in the animal category. The word dog would have no priming (facilitating) effect on the nonanimals names or pictures but should interfere with categorizing the nonanimal as such. The second graders show this interference effect.

This explanation is blatantly post hoc and should be tested empirically. However, the primary point, that children do not process words and pictures independently seems evident. That dependency, in turn, suggests that word recognition is not fully automatic and independent of the demands and context of the task by the fourth grade. The capacity to encode and recognize words independent of their particular context may be an important quality of automaticity.

Dimensions of Lexical Coding in Chinese and English

Hsuan-Chih Chen and James F. Juola
Memory and Cognition,
in press

Words can be coded in terms of their graphemic, phonemic, or semantic properties, and these three types of codes remain in long-term memory after a word is presented and read. This project was designed to test the relative strengths of these three types of memory codes both immediately after studying a list of words and 24 hours later. Further, it was decided to test for differences in memory codes between American subjects' memory for English words and Chinese subjects' memory for Chinese logographs. Because of the high degree of correspondence between phonology and orthography in English, it was expected that memory codes would be more integrated along these dimensions than for Chinese, in which there is very little systematic relation between the sound and the visual appearance of logographs.

Method

The subjects were 24 University of Kansas students who were native speakers of English and 24 natives of Taiwan. Each studied a list of words presented in his or her native language. These were 36 common nouns presented for 8 sec each. In a recognition memory test, two words at a time were presented, and the subjects had to choose which word of each pair was similar to a word on the study list. The test

pairs were blocked, such that in separate groups of trials subjects were to judge whether one of the words sounded like, looked like, or had the same meaning as one of the words on the study list. The recognition memory test was administered in two parts using different stimuli--one immediately after and one 24 hours after the study session.

Results

The results showed significant differences between Chinese and American subjects in terms of the relative strengths of memory codes. As expected, the results for English words showed a more integrated pattern, with equivalent recognition memory performance for phonemic, graphemic, and semantic information. The Chinese subjects, however, demonstrated much greater memory for visual codes of logograhps than for phonemic or semantic information.

Discussion

Differences between how words are coded and remembered in different languages have obvious implications for theories of reading. It is important to note that the initial coding of logographic characters relies heavily on purely visual processes, and phonemic analysis is apparently relatively less important than it is in encoding English words. A review of the literature indicates that despite word encoding differences, more central processes of comprehension are probably quite similar across different linguistic groups.

Implications

We are beginning to acquire some understanding of the development of automatic word recognition. Chen and Juola's work with Chinese and American adults makes it obvious that adults' integrate phonemic, graphemic, and semantic information in recognizing words. Our work, and that of others, using Stroop-like interference tasks also leads to the same conclusion. Adults show increasing amounts of interference in color and picture naming from distractors that are identical letter strings, nonwords, pseudowords, and meaningful words. There is evidence that differences in semantic content also affect the amount of interference.

Children begin to show some of these same componential effects of word recognition as they begin to learn to read; letters, nonwords, psuedowords, and words used as distractors come to interfere with the designated task of

naming pictures or words. Our research has clearly shown that interference effects emerge early, but also continue to change as reading skill increases, at least through the sixth grade or higher.

The emergence of interference in Stroop-like tasks may represent only a minimal type of automaticity. The results of Schadler and Thissen's (1981) study are evidence that minimal reading skill, the ability to name letters, is enough to cause some interference with color-naming. Similarly, the children's performance in the picture-recognition study suggests that once words enter the cognitive system, they are processed to completion, taking priority over and delaying execution of the designated task. This priority effect has not been found with adults, who seem to evidence interference only when the atuomatic process is completed before the intentional one. We suggest that, whereas adults process the designated and distractor elements of Stroop-like tasks separately, younger and less-adept readers process the meaningful information in a task domain in a more holistic fashion. They are not selective.

II. COMPREHENSION OF BRIEF PASSAGES

Intermediate to the basic task of accessing the meaning of isolated words and understanding extended passages of text is the step of integrating meaning within or between sentences. This operation is addressed in the next study. The integration process is explored by determining and comparing reading times for sentences in which an exemplar must be linked with its referent category. Although the research is written as one study containing four tasks, it is better conceptualized as consisting of four related experiments, each of which places different demands on the reader.

Forward and Backward Search Processes in Sentence Integration

Frances J. Friedrich

Much of the cognitive activity that occurs during reading involves the integration of semantic information. Through integration processes, relationships between sentences are recognized or established, and individual pieces of information are organized. Comprehension of a text depends to a great extent on the various coherence markers built into the text (e.g., Clark, 1977; Halliday & Hasan, 1976). Specific linguistic cues may be used to identify the relationship between individual propositions. For instance, a connector such as "and" signals a parallel relationship between two elements of a text. Definite articles such as "the" often indicate that the subsequent noun is a specific instance of a previously established concept. Pronouns provide even more explicit referential cues; in order for a pronoun in text to be correctly interpreted, an appropriate antecendent must be identified.

The present study was concerned with the on-going processes involved in the integration of information across sentences. To a large extent, work in this area has focused on syntactic factors influencing the identification of antecedent-anaphor relations in text. The relationship between antecedent and anaphor phrases is that of a shared reference; the antecedent establishes an identity and the anaphor refers to it through repetition or substitution. Comprehension of the connection between them enables

information from both phrases to be integrated and stored as a single unit in memory.

The effect of inferential processes on integration time has been demonstrated in several studies that varied the nature of the antecedent-anaphor relationship (e.g., Yekovich & Walker, 1978; Frederiksen, 1979; Garrod & Sanford; 1977). Garrod and Sanford found that the strength of the antecedent- anaphor relationship affected comprehension times for the anaphor sentences, with the sentences containing high frequency exemplars producing shorter comprehension times than those containing low frequency exemplars. Frederiksen (1979) found that reading times for target sentences were fastest when the appropriate referent was the subject of a prior sentence. If a sentence intervened between the referent and the target pronoun, and another pronoun sharing the same referent occurred in the predicate of the intervening sentence, the topical status of the referent was reduced. Presumably the noun phrase that served as the subject of the intervening sentence gained in topical status at the expense of the true referent.

These findings raise some questions about the nature of the antecedent identification process. For instance, are syntactic cues needed to initiate a search and identification process, and what is the role of the semantic component of text in the integration process? Carpenter & Just distinguished two ways by which verb-agent integration could occur. According to a forward inference model, the case slots associated with the verb in the introductory sentence are activated as the verb is processed, facilitating the link with the subsequently presented agent. Thus, a verb such as "to murder" will initiate a search through subsequent sentences for information relevant to the agent or the instrument, while the verb "to die" will not. In the backward inference model, on the other hand, the search for a referent begins when the agent sentence is read. Information stored in memory is scanned until an appropriate referent is found or established through inference. Facilitation for certain verb-agent matches occurs because the verbs with case entailments are already marked to receive additional information.

Although their study was not specifically designed to evaluate search processes, Carpenter & Just argued that support for the backward inference process was found in the form of a serial position effect for agent sentences. They contended that decision times were longer for sentences that occurred later in the paragraph because a more complex memory structure had to be searched before information in the agent sentence could be integrated. However, certain features of Carpenter & Just's study may have encouraged the

use of a memory search that was not essential to the integration process. For example, the consistency judgment task required subjects to store and check specific details for conflicting information. The integration process itself could have been completed before a backward search for information necessary to the consistency judgment was initiated.

Garrod & Sanford (1978) attempted to test the roles of syntactic and semantic information more directly. By varying the presence of semantic and syntactic cues in anaphoric sentences, they were able to demonstrate that an antecedent checking process occurred whenever semantic cues were available, regardless of the presence or absence of syntactic cues. They argued that a semantically- driven forward search can account for this phenomenon but a syntactically- driven backward search cannot.

Garrod & Sanford suggested that the continuity of the text might determine whether a forward or backward search process is engaged. According to this view, the semantic information in the first sentence of a text evokes a particular thematic frame in memory; the attached terminal nodes stay "open" and ready to receive additional information. If the topic of the second sentence is different, a new frame and set of terminals is opened and the first set is closed. If information in subsequent sentences has a referent in a closed location, a search through memory is necessary in order to find and re-open the appropriate memory structure.

At this point it is unclear whether sentence integration processes involve forward or backward search processes, or both. The effects of topicality and text continuity on integration have not been adequately investigated in the past. One difficulty in resolving the question on the basis of past research is that the nature of the anaphoric relations and the task demands vary greatly from study to study. The goals of the reader have been shown to affect both the coding of information within sentences (e.g. Aaronson, 1976) and macrostructure processing, including integration of information across sentences (e.g., Graesser, Hoffman & Clark, 1980). In the evaluation of search processes, the nature of the task demands may be an important factor, particularly if both forward and backward search mechanisms are available. As noted earlier, the factual consistency judgment task used by Carpenter & Just (1977) would seem to require some backward search in order for a correct decision to be made. It is unclear whether the evidence of a backward search that they found would also be obtained under less demanding comprehension conditions. In order to develop an accurate

account of the nature of on-going integration processes, it is important to determine how these processes are affected by specific task demands.

Statement of the Problem

The present study was designed to test Garrod & Sanford's (1978) model of the sentence integration processes involved in reading. If topical continuity in a text is an important factor in sentence integration, then the nature of the semantic information in a sentence that intervenes between two related sentences should determine whether a forward or backward inference process is employed. If, on the other hand, integration always involves a backward inference, as Carpenter & Just suggested, a manipulation of the focus of the intervening sentence should have little effect. Instead there should be a serial position effect for the position of the anaphoric sentence in the paragraph. Processing time should always be longer for the third sentence than for the second because a more complex memory structure has to be searched before the new information can be integrated.

The specific integration process investigated consisted of the identification of an antecedent-anaphor relation within short paragraphs. Prior research has indicated that manipulation of the anaphoric relation allows relative integration time to be inferred from reading times for the anaphoric sentences. In the present study, semantic category labels served as antecedents and two types of category members, representing high or low production frequency, were used as anaphors. Because a category name is more closely associated with a high frequency exemplar than a low frequency exemplar, the inference process involved in the antecedent-anaphor identification should be easier and more rapid for the high frequency exemplar. Thus, differences in response times for sentences with high vs. low frequency exemplars would indicate that antecedent-anaphor identification has occurred.

The topical continuity, or context, of the paragraph was varied in two ways. One factor was the position of the anaphor sentence in the paragraph. When the anaphor sentence occurred as the second sentence, it was contiguous with the antecedent sentence; when it occurred in the third position, a filler sentence providing additional context intervened between the antecedent and the anaphor. The second factor affecting text continuity was the nature of the filler sentence, which was designed either to maintain the emphasis of the paragraph on the category topic or to shift the focus to another aspect of the antecedent sentence.

The nature of the reaction time task used to measure integration processes is an important consideration for two reasons. First, the individual reader can adapt both reading speed and level of comprehension to specific goals and strategies. It is possible that the nature of the integration processes may be altered accordingly. Second, integration time must be inferred from the overall response time to an individual sentence. Certain types of task demands may affect decision time in such a way that the nature of the integration process is obscured. To evaluate the effects of task demands on integration processes, antecedent-anaphor identification time was measured in the context of four different task situations. In the first task, subjects were asked to judge the meaningfulness of individual sentences in each paragraph; thus, no integration with or reference to prior information was necessary to make the decisions. Evidence that integration occurs under these instructions would indicate that integration processes proceed automatically, independent of specific task demands. The second task was a thematic consistency task in which subjects were asked to determine whether or not a sentence was consistent with the theme established by prior sentences. This decision should require the integration of general information but should not necessitate a storage and search of details. The third task required a factual consistency judgment, and therefore made greater demands upon memory than the other tasks. In order to determine whether details in a sentence contradicted prior information, subjects had to store and compare specific facts from each sentence. Carpenter & Just (1977) used a similar task in their study of integration processes and found some evidence to support a backward search model. The factual consistency task was included in the present study in order to determine the extent to which memory demands affect integration processes. In the fourth task, no decision as such was necessary; subjects were instructed to read for comprehension, and the reading time for each sentence was recorded. Because reading speed and level of comprehension appear to be sensitive to various goals and strategies, it is likely that any decision task imposed upon the reader will affect the nature of the on-going cognitive processes. The inclusion of a simple reading task is important in order to evaluate integration processes in a reading comprehension situation that is free of explicit decision requirements.

Method

Design. Integration was investigated through a manipulation of antecedent-anaphor relationships in short paragraphs. The antecedents were category names and the anaphors were exemplars from these categories. Each

paragraph consisted of three sentences, one that contained the antecedent, one that contained the anaphor, and a filler that provided additional information about a topic mentioned in the antecedent sentence. The nature of the paragraph was varied along three dimensions: exemplar frequency (high vs. low), anaphor sentence position in the paragraph (contiguous with the antecedent sentence vs. noncontiguous), and topical focus of the filler sentence (same-focus vs. different-focus). These factors were completely crossed, producing a 2 X 2 X 2 factorial and a total of eight paragraph conditions.

Paragraphs generated in this manner were presented in four different instruction situations. The four response time measures consisted of judgments of sentence meaningfulness, thematic consistency, factual consistency, or simple comprehension time. The data from each task were analyzed separately because decision difficulty could not be equated across the different tasks.

Subjects. Subjects were undergraduate volunteers who received either course credit or $3.00 payment for participation. All were native English speakers, with normal or corrected-to-normal vision. A different group of subjects participated in each of the four tasks.

Some subjects failed to correctly detect the negative instances in the decision tasks. A high error rate of this kind suggests that these subjects had either misunderstood the instructions or were not reading the sentences carefully. Therefore, an error criterion based on undetected negative instances was established. The data of subjects who failed to detect 35 per cent or more of the negative instances were not used, and additional subjects were tested in the appropriate decision task.

Data analysis was based on the data of 24 subjects in each task. A total of 28 subjects participated in the sentence meaningfulness task, but the data from four subjects were not included in the analysis due to error rates that exceeded the error criterion. In the thematic consistency task, the data from four of the 28 subjects were not used for the following reasons: two subjects exceeded the error criterion, one was disturbed by loud noise outside the laboratory during the experimental session, and one admitted to not reading the paragraphs completely in the interest of speed. Of the 34 subjects who participated in the factual consistency task, the data from 10 subjects were removed on the basis of the error criterion.

Stimulus materials. A set of 48 scenarios was developed to serve as the basis for paragraph construction. These scenarios were designed to incorporate references to specific semantic categories. For instance, the category "reading materials" was included in a scenario set in a doctor's waiting room. The 48 categories and two exemplars from each category were selected from the Battig & Montague (1969) production frequency norms, based on a sample of 442 subjects. The exemplars were chosen on the basis of frequency such that one category member represented a high frequency exemplar and the other a low frequency exemplar. The mean frequencies for the high and low frequency exemplars were 286 and 14, respectively, and the minimum frequency difference per pair was 74. A t-test of the average production frequency for the two sets of exemplars confirmed the difference, $t(47)=16.6$, $p<.001$. Each pair of exemplars was matched for number of syllables and was matched as closely as possible for printed frequency of occurrence (Kucera & Francis, 1967). A t-test showed that the two sets of exemplars did not differ in word frequency, $t(47)=.73$, $p<.5$.

For each scenario a set of sentences was constructed according to the demands of the experimental manipulations. In the antecedent sentence, the category name was used as the subject of the sentence and was preceded by an indefinite article. The category name served as the subject of the sentence in order to establish it as the primary topic. The anaphor sentence was then constructed using an exemplar as the subject of the sentence. The exemplar was preceded by the definite article "the", which served as a syntactic cue to indicate that the noun phrase represented a specific item that had been referred to previously. The anaphor sentence was constructed so that both the high and low frequency exemplars would be meaningful in that context. Sample sentences can be found in Table 1.

Two filler sentences were then constructed for each scenario. One was designed to maintain the focus of the paragraph on the category topic. Neither the category name, pronouns, nor other direct representations of the category were used in the filler to insure that information in this sentence could not be used as an antecedent to the anaphor sentence. The second filler sentence was designed to shift the focus of the sentence away from the category and to another topic mentioned in the predicate of the antecedent sentence. Thus, the different-focus filler was thematically consistent with the antecedent sentence but emphasized a topic other than the category name.

Table 1

Sample stimulus materials

Category: Precious stones
Antecedent sentence Some precious stones were kept in a small locked box in the study.
High frequency anaphor sentence The ruby was discovered missing later that evening.
Low frequency anaphor sentence The garnet was discovered missing later that evening.
Same-focus filler sentence Nothing of value had escaped the eye of the clever thief.
Different-focus filler sentence The study was nicely decorated, but always looked rather cluttered.
Sentence meaningfulness negatives Nothing of value had escaped the eye of the clever gardens.
The study was nicely decorated, but always looked rather gardens.
Thematic consistency negatives The vitamin content may vary depending on the quality of the soil.
Modern houses often don't have the space for lawns and gardens.
Factual consistency negative The thief removed the stones from the safe in the bedroom.

All of the sentences were less than 80 characters long, so that each sentence could be presented on a single line. There was an average of 65 characters (range= 44-77) and 11 words (range= 7-15) per sentence. An effort was made to construct fairly complex and varied sentences in order to keep the paragraphs interesting and as much like normal text as possible.

Paragraphs for the eight different conditions were generated by selecting and ordering three of these five sentences in specific ways. The category or antecedent sentence appeared as the first sentence of the paragraph in all conditions. In the contiguous conditions, the anaphor sentence, with either a high or low freqency exemplar, occurred as the second sentence in the paragraph, and the filler appeared third. In the noncontiguous conditions, the filler sentence was second and the anaphor sentence was third. The topic-maintaining filler sentence was used for all same-focus conditions, while the topic-shift filler was used in the different-focus conditions. Examples of each paragraph type are found in Table 2.

For each scenario, a set of sentences was also constructed to serve as negative instances for the various tasks. The negative sentence replaced the filler sentence of the paragraph in all cases. A negative instance for the factual consistency task consisted of a sentence that maintained the theme of the paragraph but contradicted one or more details mentioned in the antecedent sentence. For the thematic consistency task, the two filler sentences of a different paragraph were used as negatives. In the sentence meaningfulness task, the final word of the filler sentence was replaced by a word that made the sentence semantically incongruous.

In addition to these 48 test paragraphs, a set of 10 neutral paragraphs were constructed. These paragraphs consisted of three sentences but did not contain an embedded category-exemplar relationship. They were included in the experimental session so that the structure of the test paragraphs would not become predictable or monotonous.

Eight stimulus sequences were generated so that a paragraph would occur in every condition but no subject would see the same paragraph twice. Paragraphs were randomly assigned to conditions in each sequence, with the constraint that each sequence have an equal number of paragraphs in all conditions. The order of presentation of paragraphs within a sequence was randomly determined. An equal number of subjects received each of the eight sequences to insure that each scenario occurred in each condition an equal number of times across all subjects. For

Table 2

Sample paragraphs for positive trials*

High frequency/Contiguous/Same-focus

Some precious stones were kept in a small locked box in the study.
The ruby was discovered missing later that evening.
Nothing of value had escaped the eye of the clever thief.

High frequency/Noncontiguous/Same-focus

Some precious stones were kept in a small locked box in the study.
Nothing of value had escaped the eye of the clever thief.
The ruby was discovered missing later that evening.

High frequency/Contiguous/Different-focus

Some precious stones were kept in a small locked box in the study.
The ruby was discovered missing later that evening.
The study was nicely decorated, but always looked rather cluttered.

High frequency/Noncontiguous/Different-focus

Some precious stones were kept in a small locked box in the study.
The study was nicely decorated, but always looked rather cluttered.
The ruby was discovered missing later that evening.

Low frequency/Contiguous/Same-focus

Some precious stones were kept in a small locked box in the study.
The garnet was discovered missing later that evening.
Nothing of value had escaped the eye of the clever thief.

Low frequency/Noncontiguous/Same-focus

Some precious stones were kept in a small locked box in the study.
Nothing of value had escaped the eye of the clever thief.
The garnet was discovered missing later that evening.

Low frequency/Contiguous/Different-focus

Some precious stones were kept in a small locked box in the study.
The garnet was discovered missing later that evening.
The study was nicely decorated, but always looked rather cluttered.

Low frequency/Noncontiguous/Different-focus

Some precious stones were kept in a small locked box in the study.
The study was nicely decorated, but always looked rather cluttered.
The garnet was discovered missing later that evening.

*Negative trials were created by replacing the filler sentence of the paragraph with the negative instance approriate to the task.

the three decision tasks (Tasks 1, 2, and 3), one-fifth of the paragraphs in each sequence were selected to serve as negative trials. There were an equal number of negative trials in each condition.

Procedure. Each subject was tested individually and began the session by reading a set of instructions appropriate to one of the four tasks. (Complete instructions are provided in Appendices B-F). The experimenter then gave a brief oral summary of the instructions and answered any questions. The subject was seated in a semi-darkened room, approximately .65 meters from a CRT screen. At this distance the horizontal visual angle subtended by a four-letter word was about 1.2 degrees. The experimental session consisted of 10 practice trials and 50 test trials and lasted about 40 minutes. All conditions, including some negative trials, were represented in the practice series. The experimenter remained in the room with the subject during the practice series to provide feedback and answer questions but left the room during the test trials.

The stimulus materials were presented in normal case on a Teleray CRT screen driven by a PDP 11/03 minicomputer. Response times for each sentence were recorded by the computer, beginning with the onset of the sentence and terminating as soon as a response key was pressed. The response type (Yes/No) was recorded, and response times were measured to the nearest millisecond.

Subjects were instructed to keep their hands on the response keys at all times and to respond as quickly as possible. Subjects used their dominant hands for "Yes" responses and the nondominant hands for "No" responses. The sequence of events for a single trial was as follows. The trial began when the word "READY" was presented for 500 msec at the beginning of a line in the center of the screen. This warning signal served both to prepare the subject and as a fixation point for subsequent sentences. After an interval of 500 msec the first sentence of the paragraph appeared and remained on the screen until one of the response keys was pressed. At this point, the first sentence disappeared and was instantly replaced by the second sentence, which appeared at the same location. When a response was again made, the second sentence was replaced by the final sentence of the paragraph. The response to the third sentence cleared the screen. After a two-second interval, a "READY" signal appeared and a new trial began.

Instructions for the individual tasks were as follows. Subjects in the sentence meaningfulness task were instructed to simply read the first sentence of the paragraph and to press the "Yes" key when they had finished. No decision was

made on the first sentence in order to make the procedure consistent with that of the other decision tasks. For the second and third sentences of each paragraph, subjects were instructed to press the "Yes" key if the word string formed a semantically meaningful sentence and to press the "No" key if it did not.

In the thematic consistency task, subjects read the first sentence and pressed the "Yes" key when they had finished. This category sentence established the theme of the paragraph. For the second and third sentences, subjects were instructed to press the "Yes" key if the sentence was thematically consistent with prior sentences and to press the "No" key if the sentence was clearly concerned with another theme. If the subject responded "No" to the second sentence, he was instructed to judge the third sentence in relation to the thematic information in the first sentence. The instructions for the factual consistency task were identical to these, except that subjects were asked to press the "Yes" key if the sentence was factually consistent with earlier sentences and to press the "No" key if the details of the sentence contradicted prior information in any way.

In the reading task, no decision was required, and subjects were instructed to press the "Yes" key when they had finished reading each sentence. To insure that subjects actually read the material, they were informed that a recognition test would follow presentation of all of the paragraphs. They were advised not to try to memorize the sentences, but to simply read at their normal rate.

Results and Discussion

The data of primary interest were the response times for the anaphor sentences in the positive trials. The paragraphs were constructed according to a 2 X 2 X 2 factorial design, with the manipulation of three factors: exemplar frequency, anaphor sentence position, and filler sentence focus. However, the response time for the anaphor sentence cannot be affected by the filler sentence manipulation when the anaphor sentence is contiguous with the antecedent sentence and the filler sentence is in the third position. Therefore, the two contiguous conditions (same-focus and different-focus) were collapsed into a single condition. The result is a single factor of context type with three levels: contiguous, noncontiguous/ same-focus, and noncontiguous/different-focus. The factors of exemplar frequency and context type were tested in separate analyses for each task.

In all of the analyses, both exemplar frequency and context type were treated as fixed factors and subjects was treated as the sole random factor. It has been argued (Clark, 1973) that factors representing a manipulation of language stimuli should be treated as random factors in order to allow generalization to all language materials. However, the use of a random-effects model is based on the assumption that the levels of the language factor are selected at random (Wike & Church, 1976). In the present study, selection of materials was highly constrained in order to provide the controls necessary to test the hypotheses in question. Therefore, an analysis of variance using a fixed-factor model was considered to be the most appropriate method of analysis.

Task 1: Sentence Meaningfulness. The mean response times for sentence meaningfulness judgments (see Figure 1) were analyzed in a 2 (exemplar frequency) X 3 (context type) repeated-measures analysis of variance. The two main effects were found to be significant: exemplar frequency, $F(1,23)= 8.79$, $p<.007$, and context type, $F(2,46)= 4.23$, $p<.02$. The main effect for exemplar frequency reflects the fact that response times for low frequency exemplars were longer than response times for the high frequency exemplars ($X= 4.90$ sec and 4.43 sec, respectively). This finding indicates that the relationship between the category name and the category exemplar did affect response time for the anaphoric sentence, even though the identification of the antecedent-anaphor relationship was not a necessary part of the decision process in the meaningfulness task.

The main effect of context type provides additional information about the nature of the integration process. Compared to the contiguous condition ($X= 4.59$ sec), in which the anaphor sentence immediately followed the category sentence, the response times for the noncontiguous/ same-focus condition are shorter ($X= 4.44$ sec) and the response times for the noncontiguous/different-focus condition are much longer ($X= 4.97$ sec). Multiple comparisons of the means using a protected t-test (Welkowitz, Ewen, & Cohen, 1976) show that the different- focus mean is reliably different ($p<.05$) from the other two means ($t(45) = 3.3$ and $t(45) = 2.3$), but that the decrease in response time between the contiguous and the same-focus conditions is not reliable. It is clear from these results that increased distance between the antecedent and anaphor per se does not make integration more difficult. Rather, the nature of the intervening sentence is an important factor in determining the speed with which an antecedent-anaphor relationship can be identified; the integration process is slowed only when the topic of the intervening sentence is unexpected.

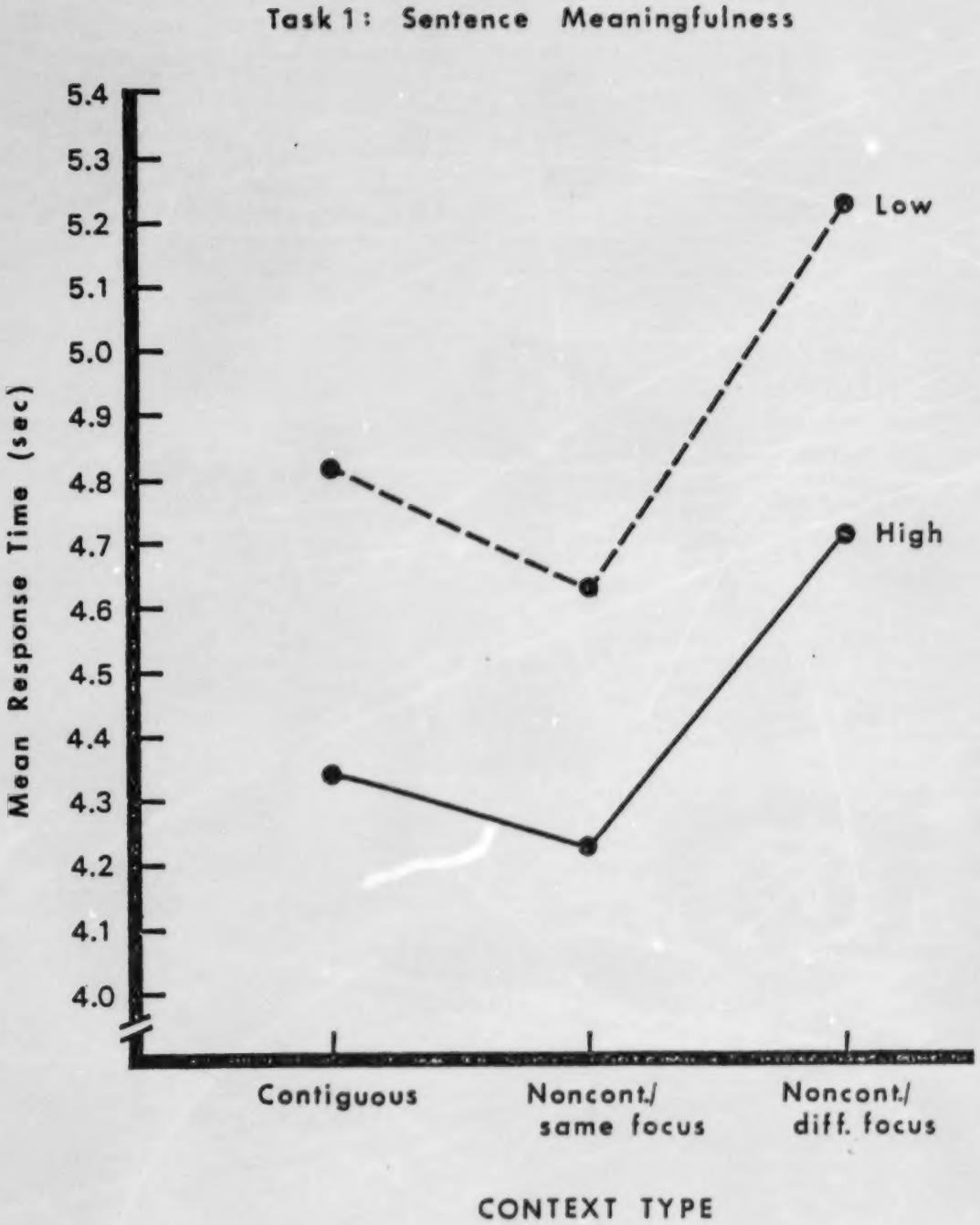

Figure 1. Mean response time in seconds for anaphor sentences as a function of exemplar frequency and context type.

Errors occurring on the anaphoric sentences were also analyzed. Because the inconsistency was always located in the filler sentence, all anaphor sentences were meaningful. Therefore, an error occurred when subjects judged the anaphor sentence to be meaningless. The mean errror proportion for the anaphor sentences was .019. A 2 X 3 repeated-measures analysis of variance was performed on the arcsine-transformed error proportions; there were no significant effects of any factors. Thus, the accuracy of the meaningfulness decision was not affected by either exemplar frequency or context type. This finding is in keeping with the nature of the task, since the decision depended upon information within each individual sentence.

The results of the sentence meaningfulness task are interesting in terms of an evaluation of sentence integration processes. According to the forward search model, the cognitive system expects additional information about the current topic. This view suggests that 1) integration of information among individual sentences occurs automatically, and 2) the topical focus of material intervening between the antecedent and anaphor should influence the antecedent identification process. The backward search model, on the other hand, predicts that the distance between the antecedent and the anaphor is an important factor, such that any intervening material should slow down the integration process. The results of the sentence meaningfulness task support the forward search model in several respects. First, the exemplar frequency main effect indicates that integration occurred in the form of an identification of the antecedent-anaphor relationship in the paragraph. As this identification was not a necessary part of the decision process in the meaningfulness judgment, it appears that integration occurred automatically. Second, although the distance between the antecedent and anaphor did not affect decision time, the nature of the intervening material, in terms of the topical focus of the filler sentence, did prove to be a crucial factor in the antecedent identification process.

It could be argued that syntactic cues, in this case the use of a definite article before the anaphor, function to trigger an automatic backward search for the antecedent. Thus, evidence of automatic integration in a task that demands comprehension only at the sentence level may not be sufficient to distinguish between forward and backward search processes. Nevertheless, a backward search model cannot account for the effects of the context type manipulation, specifically the dramatic increase in response times for the different-focus condition with no accompanying increase in the same-focus condition. Even if the backward search does occur automatically, the distance between the

antecedent and anaphor, rather than the nature of the intervening material, should be the main predictor of response times. This was clearly not the case since the same-focus condition actually produced a slight decrease in response times relative to the contiguous condition.

Task 2: Thematic Consistency. The mean response times for the judgments of thematic consistency are presented in Figure 2. A 2 X 3 repeated-measures analysis of variance revealed a significant main effect for exemplar frequency, $F(1,23)= 8.10$, $p<.009$. The main effect for context type was marginal, $F(2,46)= 2.91$, $p<.065$, but the interaction of exemplar frequency and context type did reach significance, $F(2,46)= 3.27$, $p<.047$.

Response times for the high frequency exemplar sentences were shorter than those of the low frequency sentences (X= 3.10 sec and 3.44 sec, respectively). However, the pattern of means across context type was strikingly different for the two levels of exemplar frequency. The means of the high frequency conditions were virtually identical for the contiguous (X= 3.06 sec), same-focus (X= 3.14 sec), and different-focus (X=3.10 sec) contexts. This was confirmed by t-tests comparing the means of the three conditions, none of which reached significance. The low frequency exemplar sentences, on the other hand, showed a pattern of means similar to that found in Task 1: a decrease in response time between the contiguous (X= 3.40 sec) and same-focus (X= 3.18 sec) conditions, and a sharp increase in response times for the different-focus condition (X= 3.73 sec). The t-tests indicated that these three means were reliably different ($p<.05$) from one another, $t(45) = 3.0$ (contiguous vs. different-focus), $t(45) = 5.0$ (same- vs. different-focus), and $t(45) = 2.0$ (contiguous vs. same-focus).

An analysis of arcsine-transformed error proportions using a 2 X 3 repeated-measures analysis of variance revealed main effects for both exemplar frequency, $F(1,23)= 7.79$, $p<.01$, and context type, $F(2,46)= 6.48$, $p<.003$. More errors were made on the low frequency exemplar sentences (X= .049) than on the high frequency sentences (X= .014). In addition, there was an increase in errors across context type, with a mean error proportion of .005 in the contiguous condition, .016 in the same -focus condition, and .073 in the different-focus condition. However, t-tests showed that while the different-focus mean differed significantly ($p<.05$) from the other two means ($t(45) = 3.33$ and $t(45) = 2.83$), the means of the contiguous and the same-focus conditions were not reliably different.

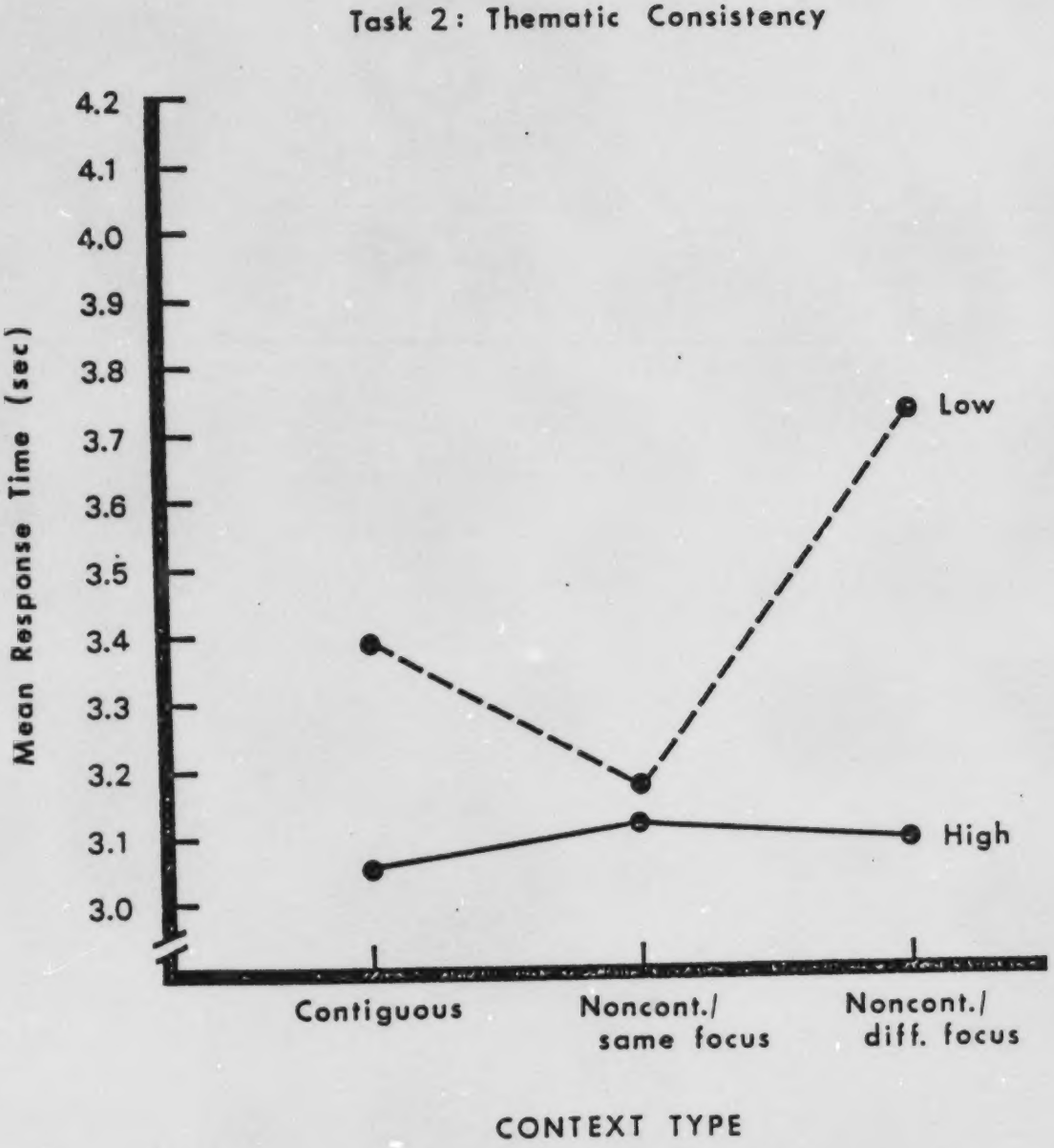

Figure 2. Mean response time in seconds for anaphor sentences as a function of exemplar frequency and context type.

The anaphor sentence response times provide a useful measure of antecedent-anaphor identification time, as reflected in the main effect of exemplar frequency. However, the identification process seems to be quite different for the two levels of exemplar frequency. For the low frequency exemplars, the integration process is facilitated by additional context, in the form of an intervening sentence, when that context provides information relevant to the category topic. If the intervening sentence shifts the focus away from the category topic, however, the antecedent-anaphor identification process is inhibited. In terms of the forward search model, what appears to be happening is that the same-focus filler sentence reinforces the likelihood that additional information relevant to the current topic will be forthcoming, thereby making the antecedent identification process easier. The different-focus filler, however, causes a different aspect of the antecedent sentence to be emphasized, such that the forward search mechanism expects additional information about a feature other than the category topic. In order to identify the appropriate antecedent for the exemplar, a backward search must be initiated. It should be remembered that the facilitation due to the additional context in the same-focus filler could not be the result of using information in the filler sentence as an antecedent. These fillers were constructed to insure that they did not contain pronouns or other direct representations of the category name. This pattern of results must therefore reflect the integration of information in the category and exemplar sentences.

For the high frequency exemplar sentences, context type had no effect on response time. It is likely that this result reflects a floor effect. In the context of a thematic consistency judgment, a decision could be made as soon as enough information was obtained to determine the topic of each sentence. The relationship between the category name and the high frequency exemplar was apparently so strong that a judgment could be made as soon as the anaphor was encountered. That is, the strength of that relationship negated the influence of the irrelevant information presented in the different-focus filler sentence.

Task 3: Factual Consistency. The mean response times for anaphoric sentences in the factual consistency task are presented in Figure 3. In the 2 X 3 repeated-measures analysis of variance, only the main effect of context type reached significance, $F(2,46)= 7.54$, $p<.001$. The mean response time for low frequency exemplar sentences was longer than for the high frequency sentences ($\bar{X}= 5.36$ sec and $\bar{X}= 4.90$ sec, respectively), but the difference was not

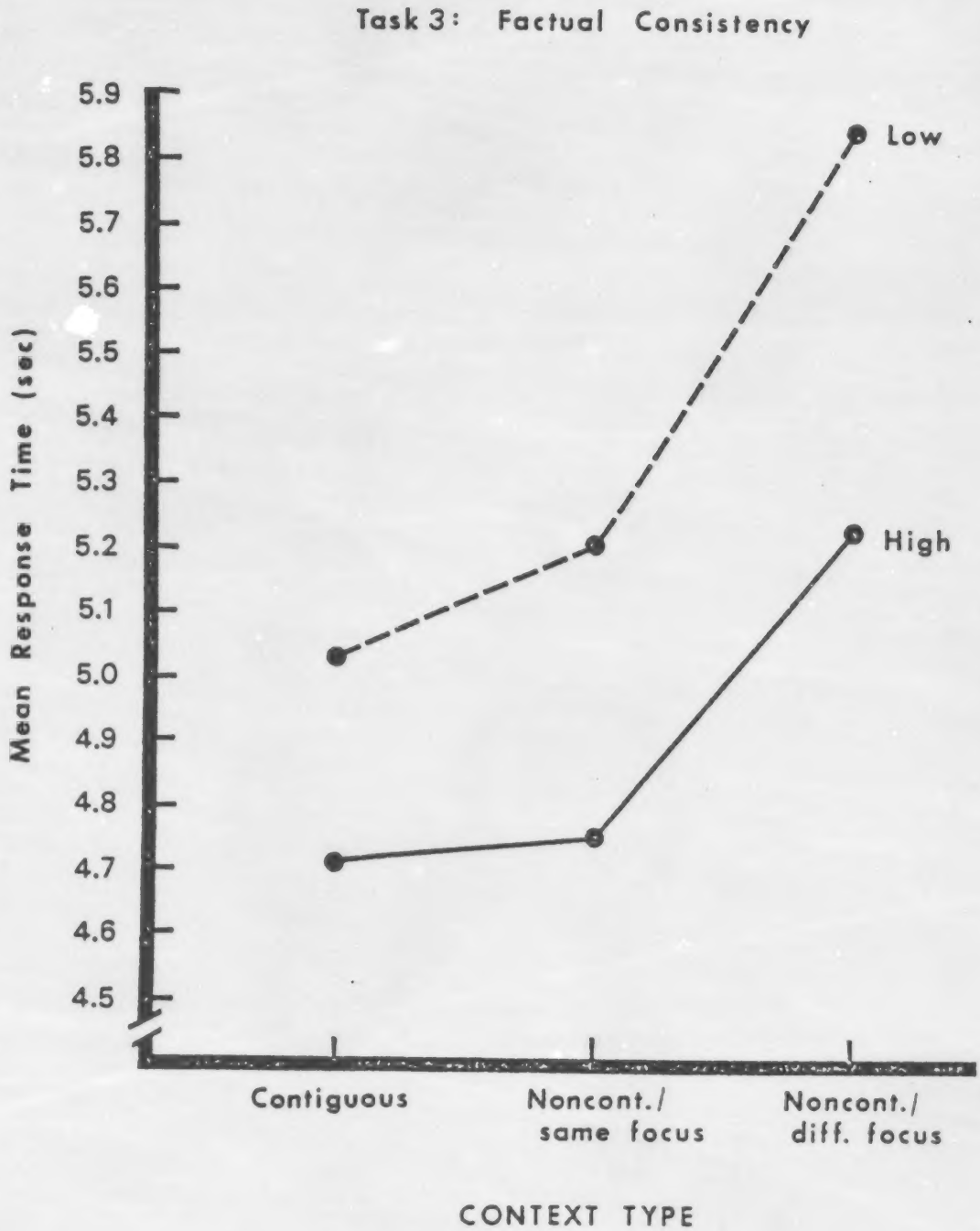

Figure 3. Mean response time in seconds for anaphor sentences as a function of exemplar frequency and context type.

reliable, $F(1,23) = 3.33$, $p<.08$.

The means of the three levels of context type present a slightly different pattern of results in the factual consistency task than in the first two tasks. In this case, there was an increase in response times from the contiguous condition (X= 4.88 sec) to the same-focus condition (X= 5.0 sec) and another increase in the different-focus condition (X= 5.54 sec). However, t-tests revealed that only the different-focus mean was reliably different from the other two conditions ($t(45) = 3.6$ and $t(45) = 3.0$).

A 2 X 3 repeated-measures analysis of arcsine-transformed error proportions showed a similar pattern of results. Again, although more errors occurred in the low frequency sentences (.082) than in the high frequency sentences (.053), only the main effect of context type was significant, $F(2,46)=5.36$, $p<.008$. The proportion of errors was lowest for the contiguous condition (.023), with a sharp increase in errors in the noncontiguous conditions (.075 for the same-focus and .104 for different- focus). On the basis of t-tests, the latter two means did not differ from one another, but the contiguous condition mean was reliably different from both ($t(45) = 1.89$ and $t(45) = 3.25$).

The factual consistency judgment was clearly more difficult than the other two types of decisions discussed previously. The difficulty was probably due to the increased memory requirements; because prior sentences were not available for reference, the information from those sentences had to be stored rather completely. Evidence of this difficulty is found in the number of subjects whose data were not considered on the basis of the undetected negative error criterion and in the relatively high error rate in the anaphor sentences (.067 overall). The difficulty of the decision also increased the variability in the reaction time data, such that a mean difference between the levels of exemplar frequency of 460 msec, which was sufficient to reach significance in the other two decision tasks, was not significant in this case. Interpretation of the response time data is therefore difficult, as the integration time appears to be obscured by other factors affecting response time.

It is important to note, however, that the additional context of the same-focus filler did not facilitate the responses in the factual consistency task as it did in the thematic consistency task. Rather, responses were both slower and less accurate when the anaphor sentence appeared third, regardless of the nature of the intervening sentence. These trends are consistent with the findings of Carpenter & Just and suggest that the heavy memory demands of the task

may have encouraged the use of a backward search process. That is, as each sentence was read, memory representations of prior sentences had to be searched to identify any contradictory information. Thus the comparison process became more difficult with an increasing amount of prior information. The fact that response times in the different-focus condition were significantly longer than those in the same-focus condition suggests that the shift in topical focus in the intervening sentence complicates the comparison process further. It is not clear, however, whether the presence of a different-focus filler results in a more complex memory structure or whether it affects the order in which the prior information is searched. In any case, the pattern of results across context type does support the contention that the backward search process becomes more difficult with increased prior context, even though we cannot assume that these response times specifically reflect the antecedent-anaphor identification process.

Task 4: Reading. Mean reading times for the anaphor sentences (see Figure 4) were analyzed in a 2 X 3 repeated-measures analysis of variance, and the main effect for exemplar frequency was the only factor to reach significance, $F(1,23)= 14.09$, $p<.001$. Sentences with high frequency exemplars produced shorter response times than the low frequency sentences ($\bar{X}= 4.28$ sec and $\bar{X}= 4.49$ sec, respectively). These results suggest that integration processes are reflected in the reading times, with the identification of the antecedent-anaphor relationship being somewhat slower for the low frequency exemplars. Although the context type main effect did not reach significance, the pattern of means is similar to that found in other tasks. For the high frequency exemplars, the means of the contiguous and same-focus conditions were virtually identical ($\bar{X}= 4.22$ sec and $\bar{X}= 4.20$ sec, respectively), but there was an increase in response time in the different-focus condition ($\bar{X}= 4.44$ sec). The means of the low frequency exemplars showed a small but steady increase across the three conditions ($\bar{X}= 4.42$ sec, 4.50 sec, and 4.56 sec).

The lack of a significant context type effect in this task is somewhat perplexing, particularly in light of the fact that context effects have been demonstrated in other studies using a simple comprehension time measure (e.g., Haviland & Clark, 1974). One possible source of difference may be the nature of the anaphoric relation, since other studies have in general used repeated-noun or synonym anaphors in very simple sentence structures. However, it is more likely that the absence of explicit decision demands allowed subjects to employ a number of different strategies, thus obscuring any single pattern. For instance, although

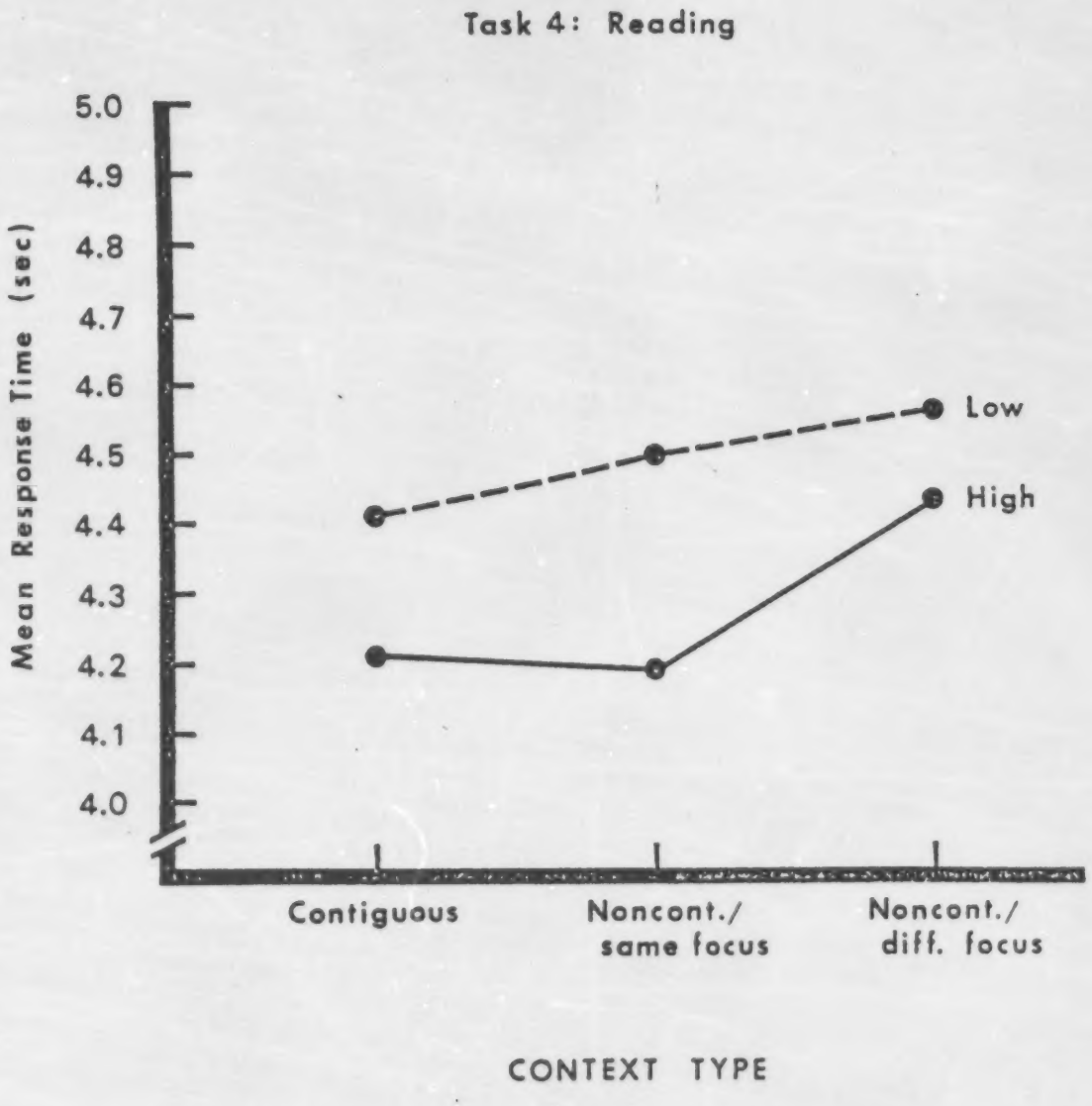

Figure 4. Mean response time in seconds for anaphor sentences as a function of exemplar frequency and context type.

they were instructed to read for comprehension, the expectation of a subsequent recognition test may have prompted some subjects to try to memorize each sentence, while others read as quickly as possible.

General Discussion

The present study speaks to three major issues concerning sentence integration processes. First, to what extent does comprehension time for an individual sentence reflect the integration of information across sentences? Reading is a composite of a number of processes, many of which (e.g., encoding, lexical access) are involved in the formation of a semantic representation of the individual sentence, independent of context. The separation of macrostructure processing from the lower level processing in response times must be substantiated before any other conclusions about integration processes can be drawn. Second, what is the role of text continuity in integration? Past research has explored various types of context effects, but text continuity has received little attention. This aspect of text structure would seem to be a key factor in determining the nature of the search process involved in antecedent -anaphor identification. Third, to what extent do task demands affect the integration process and our ability to infer such processing from response time measures? A wide range of response time measures have been used in the past, each of which has been assumed to provide an accurate estimation of integration time. This assumption was evaluated here by using the same set of stimulus materials in four different instruction conditions.

Response time as a measure of integration. The integration process of interest here was the identification of antecedent-anaphor relationships embedded in short paragraphs. Presumably other forms of integration occurred during the reading of these paragraphs as well, but the manipulation of the nature of the anaphor relation allowed one aspect of the integration process to be singled out while the others remained constant. Evidence of a consistent exemplar frequency effect, significant in three of the four tasks, provides strong support for the contention that response time for an individual sentence reflects processing time for the integration of information between sentences. Haviland & Clark (1974) argued that the antecedent identification time should reflect the amount of inferencing necessary to provide a suitable match between the antecedent and the anaphor. In the present study, response times for sentences containing low production frequency exemplars were consistently longer than those for the high production frequency sentences. Thus the relationship between the category name in one sentence and

the exemplar in a later sentence was an important aspect of the overall response time for the anaphor sentence.

A question remains, however, as to whether this type of relationship affects only integration processes, or whether it affects other processes, such as word recognition, as well. Neely (1977) and others have shown that when an exemplar is preceded by its category label, word/nonword decisions about the exemplar are facilitated. This type of priming could produce differential effects according to the likelihood of the exemplar, thus resulting in a frequency main effect. Garrod & Sanford (1977, Exp. 4) argued against this possibility; in a two-sentence integration study, they controlled for priming effects and found no evidence of facilitation. Although priming effects are not explicitly controlled in the present study, the manipulation of the anaphor sentence position does provide a test of the priming hypothesis. The effects of priming have been shown to diminish over time or with intervening material (Posner, Boies, Eichelman, & Taylor, 1969); therefore, if priming is the source of the exemplar frequency effect, the difference between high and low frequency sentences should be reduced when the anaphor sentence is not contiguous with the antecedent. This type of frequency by serial position interaction was not evident in any of the tasks. In the reading and sentence meaningfulness tasks, the effects of context type (including anaphor sentence position) were equivalent for both types of exemplars. The thematic consistency task did produce an interaction of frequency and context type, but not the sort predicted by the priming hypothesis. Response times for the high frequency exemplars were not affected by sentence position at all, and the context type effect for the low frequency exemplars was due to the nature of the intervening sentence rather than the position of the anaphor per se. Thus it appears that the response times for the anaphor sentences do provide a useful measure of the antecedent-anaphor identification process.

Text continuity and search processes. According to Garrod & Sanford (1978), text continuity should be an important factor in determining whether antecedent-anaphor identification occurs as a result of a forward or backward search process. The context type manipulation, which included the factor of topical focus of material intervening between the antecedent and anaphor, produced a significant effect on response times in the three decision tasks. Response times for the anaphor sentence were consistently longer when a different-focus filler preceded the anaphor than when the same-focus filler occupied that position.

The importance of the text continuity factor is quite clear, and in general the results suggest that the shift in topical focus interferes with an on-going forward search process. The pattern of means in both the sentence meaningfulness and the thematic consistency tasks indicate that additional context relevant to the antecedent topic actually facilitated the integration process; in terms of a forward search process, this suggests that the topical status of the antecedent was increased, reinforcing the expectation of still more information relevant to the antecedent. When the topic of the intervening sentence shifted away from the antecedent topic, a more time-consuming backward search had to be engaged to identify the antecedent.

The model of sentence integration proposed by Haviland & Clark (1974) and supported by Carpenter & Just (1977) is based on the assumption that a backward search operates under all context conditions and that this search process is initiated by syntactic cues in the anaphor sentence. If a backward search is the primary mechanism by which antecedents are identified, any sort of intervening material should make the search more difficult and produce longer response times. The only trends that suggest that a backward search may have occurred under all context conditions were found in the factual consistency task, in which responses tended to be slower and less accurate as the distance between the antecedent and anaphor increased. Even there topicality proved to be an important factor, as the difference between the same- and different-focus conditions was much greater than the difference between the contiguous and same-focus conditions.

Although the strong form of the backward search model is clearly inconsistent with the results of the present study, a modified version recently proposed by Clark & Sengul (1979) may be able to account for the context type effects found here. The original conceptualization of the search process assumed that items are stored and searched in memory according to the order in which they are encountered. In order to locate an appropriate referent, a match is first attempted with the most recent entity, and the search proceeds backwards until a match is found. From this view of the search process comes the prediction of a linear increase in response times due to the serial position of the anaphor: the further back one must search, the longer the search process and subsequent response times. Clark & Sengul have proposed an alternative model that gives "privileged status" to certain entities in memory. One group of items in memory consists of entities in the current sentence and one sentence back, while a second group contains entities mentioned two or more sentences back. The

premise of the "discontinuity" model is that the entities in the first group have a privileged place in working memory; they are "on stage" and readily available for identification purposes. Items in the second group have been pushed out of working memory and more effort is required to search for and identify these items. Thus Clark & Sengul predict a discontinuity in the search process, such that identification times for items more than one sentence back will be much longer than for those in the previous sentence.

Clark & Sengul tested the discontinuity hypothesis by constructing paragraphs consisting of a target sentence containing an anaphor and three context sentences, one of which contained the referent. The referent sentence position was varied so that it was either one, two, or three sentences away from the target. Based on target sentence comprehension time, subjects were able to identify the referent much faster when it was in sentence 3 than in sentence 2, and the latter was only slightly faster that when the referent appeared in sentence 1. The ratio of the difference between sentences 2 and 3 to the difference between sentences 2 and 1 was more than 4:1, which provides support for the discontinuity hypothesis. In subsequent studies, Clark & Sengul found evidence that it may be the last major clause rather than the last sentence that receives privileged status.

Although Clark & Sengul investigated the role of temporal order of acquisition on privileged status, they acknowledged that thematic considerations might be important too. Factors such as the topic of the discourse, the scene, and other assumed information may be granted privileged status regardless of their location in text. The results of the present study suggest that this is in fact the case -- the dramatic increase in response times occurred only when a topic shift intervened between the antecedent and the anaphor. Although Clark & Sengul did not manipulate topicality, the description of their materials indicates that each sentence introduced a slightly different topic, in a manner similar to the different-focus manipulation of the present study. Each paragraph was concerned with a particular theme but each of the context sentences focused on a different aspect of that scene, such that there was no overlap in the topics of specific sentences.

If the discontinuity model can be amended to include topical focus as an essential factor in the determination of privileged status, then the results of the present study are consistent with Clark that Sengul's conceptualization of antecedent-anaphor identification processes. It is not clear, however, whether the discontinuity model differs from a forward search model in any important respects.

Theoretically, the forward search model suggests a more active process, since certain expectations are generated about the nature of the information to follow. The discontinuity model merely suggests that certain information is "on stage" and available for reference, although Clark & Sengul do point out that new information is expected to be relevant to the current topic. Thus there may be a slight distinction between these two views, in terms of the active or passive nature of the mechanism, but it is a distinction that is, for all intents and purposes, untestable.

Task demands and integration processes. The results of the present study indicate that the nature of the decision task used to investigate on-going integration processes affects the pattern of response times for individual sentences. Each of the four tasks employed produced a different set of significant factors, reflecting differences in decision difficulty and response strategies among the tasks.

In the evaluation of specific tasks demands, it is interesting to note that context type affected anaphor sentence response time in every task except the reading task. The pattern of means in the reading task did show that anaphor sentences preceeded by different-focus fillers produced the longest response times, but this effect was not reliable. It is possible that the lack of explicit task demands allowed subjects to employ a variety of strategies and that the expectation of a subsequent recognition test may have prompted some subjects to spend more time on each sentence than they normally would have. Therefore, a task encouraging the use of a specific comprehension strategy by imposing particular decision demands may be more informative than a simple reading task in the evaluation of macrostructure process components of reading.

Among the three tasks in which a context type effect was found, the difference between the factual consistency task and the other two decision tasks is of particular importance. The factual consistency judgment was clearly the most difficult of the tasks, producing the longest response times, the highest error rates, and the greatest variability. This was the only task in which a main effect for exemplar frequency was not reliable. In addition, the factual consistency decision produced the only evidence of a serial position effect for the anaphor sentences. Error rates were significantly higher in the same-focus condition than in the contiguous condition, and response times showed a similar trend. This serial position effect is consistent with the results reported by Carpenter & Just (1977) who used a similar task and argued that their findings supported a backward search model. It is clear from the present study

that memory demands are a key factor in producing such a serial position effect. Under the less severe memory requirements of the other tasks, response times and error rates in the same-focus condition were equivalent to or less than those in the contiguous condition.

Evidence of differential decision difficulty was also found in the exemplar frequency by context type interaction in the thematic consistency task. The low frequency exemplar sentences showed a context type effect similar to that found in the sentence meaningfulness task, but the high frequency sentences showed no context type effect. As suggested earlier, the strong association between the theme established by the category topic and the high frequency exemplars appears to have negated the influence of intervening topics. When the anaphor was less closely associated with the theme of the paragraph, as in the case of the low frequency exemplars, context factors did influence response times.

Specific task demands clearly affected subjects' performance in the present study, but to what extent are basic integration processes altered by the goals of the reader? In general, it appears that various task demands do not actually influence the nature of integration processes so much as they affect how clearly integration time can be inferred from the response times measured. It appears to be the nature of the context rather than the specific task demands that determines which search process is engaged. The demands of the task, on the other hand, seem to affect how clearly the nature of the search can be evaluated. There is no reason to suspect that the integration process employed for the high frequency exemplars in the thematic consistency task is fundamentally different from that of the low frequency exemplars in the same task. It is more likely that subjects employed a different response strategy, that of responding before the entire sentence had been read. A similar argument can be made for the factual consistency task. The inclusion of a different-focus filler had the same strong effect in this task as it did in the other two decision tasks. It seems reasonable to conclude that the same sort of integration process occurred in all three decision tasks, but the integration time in the factual consistency task was obscured by the difficulty of the decision.

The choice of a particular response time measure is clearly tied to the aspect of integration being investigated. However, the results of the present study do indicate that a manipulation of the type of anaphor relation is an important feature in any evaluation of integration processes. Without such a factor, context effects are

difficult to interpret because there is no guarantee that the response times are actually reflecting integration time. The selection of a task that will provide a clear measure of integration time remains a problem, but the present study does provide ample evidence that topicality factors in text play an important role in determining the nature of the search process involved in antecedent-anaphor identification.

III. COMPREHENSION AND RAPID SERIAL PRESENTATIONS

A final area of research has been the use of a computer-controlled display to present text in a rapid, serial, visual presentation (RSVP) format. In this technique, single words or small groups of words are presented to a single locus on a cathode ray terminal (CRT) screen, and the text can be read without the need for eye movements. The main goal of this project has been to optimize the reading process at a given rate of text presentation and to compare comprehension of adult readers in the RSVP mode and the normal mode in which text is scanned with eye movements. The RSVP technique is also of interest because it allows for conrol over the input process in reading and thus allows a new method of testing theories of reading that link perceptual processes (such as eye fixation patterns) to comprehension processes.

Visual Search and Reading Rapid, Serial Presentations of Letter Strings, Words, and Text

James F. Juola, Nicklas J. Ward, and Timothy McNamara

Journal of Experimental Psychology: General,
1982, Vol. 111, No. 2

This manuscript reports the first four experiments carried out using the PDP 11/03 minicomputer and Teleray CRT display terminal. The first three experiments were reported in the final report for NIE Grant No. G-77-0010, so they will be sketched briefly here.

We were interested in the feasibility of using a computer-controlled display to study the process of reading as well as to explore some novel methods of text presentation that could lead to improved reading for comprehension. Specifically, we wanted to test the rapid, serial visual presentation (RSVP) method of text presentation. In this method, one or several words at a time are presented at a central location of the screen, and an entire text is sequentially presented at this location. Thus, the text can be read wihout eye movements, and the question to be answered is whether optimal means of RSVP presentation can result in improved reader comprehension over what could be obtained in the same amount of time if text is read normally with eye movements.

The first study tested the perceptibility of the dot-matrix characters generated on the CRT screen. The task for subjects was to detect the presence or absence of a given target letter in a string of from three to six letters, presened for 250 msec, which was preceeded and followed by a visual masking field. Performance was very good; responses averaged about 97% correct with a mean response time of about 750 msec. Further, a word superiority effect was found in that responses were faster and more accurate for displays that were common words than for those that were nonsense strings of letters.

A second study generalized the results of the first experiment to a search task using a single target letter followed by a rapid sequence of displays. The sequences included either five words, five legally spelled but meaningless pseudowords, or five illegally spelled and unpronounceable nonwords. Each letter string was displayed for 100, 200, or 300 msec, equivalent to a rate of from 600 to 200 words per minute (WPM). Performance was again very good, ranging from a mean of 80% correct in the most rapid sequential condition to about 89% correct in the slowest. Again, a word superiority effect was observed in all conditions. The results of the first two studies demonstrated that the perceptual characteristics of the display were of sufficient quality to ensure accurate letter identification in brief single presentations as well as in the RSVP mode.

A third experiment again used a search task, but the displays were RSVP sequences of nine words, presented one word at a time at rates of 50 or 100 msec per word (equivalent to 1200 or 600 WPM). The word sequences were either normal sentences or scrambled versions of those sentences. Performance varied from 76% to 96% correct across conditions, with a significant advantage for normal sentences over scrambled sentences at both presentation rates. The results demontrated that sentences could be read with some level of undestanding even at the highest rates used, and that both orthographic structure (Experiment 2) and linguistic structure (Experiment 3) could be used to improve search performance of RSVP displays. Since skilled readers presumably make use of orthographic and linguistic information, among other sources, to facilitate reading, the results are encouraging for the viability of the RSVP technique as a useful tool for the study of reading comprehension processes.

A fourth experiment was designed to be the first study of reading for comprehension in our laboratory using the RSVP method of presenting text. Since the typical skilled reader moves his or her eyes forward about 8-10 character

useful for the study of reading comprehension processes with control over the information input. It is also likely that improved RSVP conditions can be found that might result in comprehension superior to that obtained for the same text read normally. Such a result could have enormous implications for decisions about how text should best be presented on computer displays in the future.

Reading with and without Eye Movements: A Reply to Just, Carpenter, and Woolley

Nicklas J. Ward and James F. Juola

Journal of Experimental Psychology: General,
1982, Vol. 111, No. 2

This manuscript is a reply to Just, Carpenter, and Woolley (in press) who provide a critique of our earlier results (Juola, et al., in press). Their major point is that in two methods that subjects use to read text on their own (normal eye movements vs. the use of a button to expose each word successively), there are systematic variations in the relative amounts of time spent on different sections of the text. That is, certain words tend to attract longer eye fixations or exposure times because they are unusual, important, or not predictable from context. In the typical RSVP task, such variations are not incorporated into the display times. Just, et al. thus criticize the RSVP procedure as not only being unnatural but also lacking some of the characteristics of normal eye movement behavior that might be important for comprehension processes.

Our reply to Just, et al. emphasizes the fact that not all characteristics of normal eye movements might be necessary or even conducive to comprehension. Rather, RSVP conditions might result in superior reading precisely because some of the deleterious effects of typical eye movements are eliminated. We summarize the results of two studies that are not yet published (Ward, 1981; Ward, Chen, Ikenaga, & Juola, in preparation).

In Ward's dissertation study (Ward, 1981) we presented texts one word at a time in the RSVP mode and then tested people's comprehension and memory for aspects of the texts. Words were presented either for equal periods of time or for times determined from Just and Carpenter's (1980) model of eye fixation durations. The results showed no differences between comprehension in these two conditions. The implication is that although people might vary their eye

spaces every 250 msec, the RSVP conditions used were chosen to bracket these parameters of normal reading.

Method

Forty-eight University of Kansas undergraduates each read 24 paragraphs and were tested immediately after reading them with four multiple-choice comprehension questions. The materials used were 12 paragraphs and question sets from the McCall-Crabbs (1961) reading test (generally junior-high difficulty level) and 12 paragraphs and question sets from the Nelson-Denny (1976) reading test (generally senior-high difficulty level.

Half of the paragraphs at each level were presented in the RSVP mode and half were presented as normal "pages" of text, nearly filling the CRT screen with words. The RSVP conditions combined three "window" sizes (5, 10, or 15 characters, on the average, were presented at once) with two window durations (200 msec or 300 msec). These conditions produced six different presentation rates ranging from about 200 WPM to about 700 WPM. In the page condition, the entire paragraph was displayed for various limiting times to force the subjects to read the text at corresponding rates. Subjects were cued before each paragraph to indicate how rapidly they would have to read in order to finish the paragraph in the allotted time.

Results

Performance was measured by the mean correct responses on the four multiple-choice questions given immediately after each paragraph was read. This comprehension score varied from about 70% correct in the 200 WPM condition to about 45% correct at the 700 WPM rate. Throughout this range, performance was generally equivalent for the RSVP (or window) condition and the page (or normal reading) condition. Similarly, performance was about 15% better for the easier (McCall-Crabbs) paragraphs than for the harder (Nelson-Denny) paragraphs in both display conditions.

Discussion

The fourth experiment in this series demonstrated the viability of the RSVP format as an alternative to the normal method of reading with dynamic eye movements scanning a static page of text. Reading comprehension was shown to depend on reading rate and text difficulty in similar ways for normal and RSVP reading. From these results, we concluded that reading comprehension processes are very similar for text presented in the RSVP mode vs. text sampled with typical eye movement patterns. The RSVP method thus is

movement patterns and fixation durations in systematic ways while viewing a text, these variations are of no great importance for comprehension and need not necessarily be mimicked by a text presentation scheme that takes input control away from a reader.

A second study (Ward, Chen, Ikenaga, and Juola, in preparation) mimicked a second aspect of normal reading--the fact that some useful information about words is picked up in peripheral vision one or two fixations previous to the word's appearance in foveal vision. Thus every word is viewed several times in successive fixations in normal reading. We simulated these multiple looks in the RSVP mode by having the rightmost 1/3 or 2/3 of each window appear in the center or on the left side of following windows. The results clearly demonstrated inferior comprehension for the overlapping windows condition relative to a non-overlap window condition in which the text was presented at an equal rate.

Thus we felt that optimal RSVP conditions might not have to be closely patterned after typical eye movement patterns. Rather, our research is continuing to explore novel text presentation conditions in an unconstrained way in order to optimize the readability of text presented on a computer.

Comprehension and Memory in Reading Rapid, Serial Presentations of Text

Hsuan-Chih Chen

In this recently completed dissertation study, Chen compared the effects of different text structures and reader-skill levels on paragraph comprehension. The paragraphs were presented in a normal page format on a CRT screen or as rapid sequences of two or three words at a time (the RSVP program was set to show an average of 12 characters at a time on the computer terminal's screen).

Method

Before reading a series of pararaphs in page and RSVP formats, Chen gave his subjects (students at the University of Kansas) several reading tests. He measured their reading speed and comprehension levels using the Nelson-Denny (1976) text. Each subject was given a reading efficiency score, based on the product of their WPM reading rate and mean

percent correct for the comprehension questions on the Nelson-Denny test. A reading span measure for each subject was obtained following the procedures of Daneman and Carpenter (1980). Reading span was defined as the number of final words from a series of sentences that could be recalled after orally reading the sentences. Subjects read a number of sentences sets and then attempted to recall the last word from each sentence in the set in order. A modified ascending method of limits was used to estimate the span for each subject (range = 2 to 6).

In a second session, subjects read a series of paragraphs in either page or RSVP format and at either normal or forced, faster-than-normal rates. Half the paragraphs were presented exactly as written, and half were shown with the sentences scrambled out of order. Comprehension was assessed by a sentence rating task. After reading each paragraph, the subjects read a series of sentences containing words from the paragraphs. These test sentences were (1) exact copies of sentences, (2) true paraphrases, or (3) false (contradictory) paraphrases of sentences in the paragraph. The subject rated each sentence on a scale of 1 (sure it was not in the paragraph) to 4 (sure it was in the paragraph). The sentence rating task was repeated at the end of the session with a new set of sentences spanning all pararaphs used in the study.

Results and discussion

The sentence rating data were separated into two measures: memory for surface structure and memory for meaning. Surface memory was defined as the difference between ratings for verbatim sentences and true paraphrases, and memory for meaning was defined as the difference between the ratings for true sentences (average verbatim and true paraphrases) and those for false paraphrases.

Memory for surface structure and meaning were correlated with each subject's reading ability as measred by WPM rate, comprehension, efficiency, and span. The reading span measure correlated most highly with performance in the sentence rating task. Therefore, subjects were divided into a high-span group and a low-span group for further analyses.

Memory for surface structure was very poor in all conditions; only the high-span subjects retained a measurable amount of vebatim information, and then only in the immediate rating test. Memory for sentence meaning was better than that for exact wording, and it showed an interesting interaction with presentation format and reader ability. High-span subjects showed little difference in ability to remember meaning between RSVP and page formats.

Low-span subjects, however, demonstrated a definite advantage in the RSVP condition. Although text structure effects were found, there were no important interactions between paragraph structure and the other variables manipulated; performance was generally worse for the scrambled paragrahs.

The results replicated our earlier work by showing no differences overall between reading comprehension and presentation format (normal pages vs. RSVP). The ability by format interaction in the present study indicates that individual differences might exist for the usefulness of the RSVP method of text presentation. Skilled readers can apparently cope with a variety of text presentation formats and comprehend equally well in most cases. Less-skilled readers by definition have more difficulty in reading pages of text and might be assisted by the RSVP method. This result has important implications for RSVP applications in remedial reading situations as well as in the teaching of reading per se.

IV. SUMMARY AND CONCLUSIONS

Our work on the development of automatic word recognition is consistent with the literature on this problem that finds evidence for early emergence of automatic processing. Others (e.g., Ehri & Wilce, 1979; Stanovich et al., 1981) have argued that speed of processing continues to develop even after the process has been automated. We suggest, however, that the process itself continues to change, becoming more context independent and more flexible as reading skill develops. We have found evidence for both increases and decreases in interference. While the emergence of interference is usually taken as evidence of automaticity, decrease in interference may actually be better evidence that the smoothly executed "thought-less" or unconscious operation of the skilled performer has really developed.

Friedrich's semantic integration tasks were used only with adults, however, the techniques seem appropriate for children, and we plan to try them in future research. The use of two or more reading purposes (normal reading, consistency) should provide us with the opportunity to identify strategy shifts by children as the skill develops and with task demands. Semantic integration of the type investigated here may become automatic, partially or fully, as reading skill develops. Forward and backward searches for anaphoric relationships would presumably be executed automatically if the terms involved are not unreasonably separated. Study of the development of such a skill should facilitate our understanding of components of the reading process.

Our work on RSVP reading has encouraged us to seek new and improved ways to present text using a computer. As computer terminals become more common in our lives and are used more often as sources information, it becomes imperative to determine optimal ways of displaying information on computer screens. We believe that there are ways, such as RSVP, to present text that may be superior to the traditional page-by-page display and reading method; i.e., static pages scanned with eye movements. If so, these results could have broad applications in both professional and educational environments, as computers are used to increase public accessibility to information of all kinds.

There is no doubt that understanding the process called reading and its development is a large and complex task that is just beginning to show progress. Furthermore, the problem is not a static one but is changing as communication technology changes. The current generation of readers grew

up with television as a new visual medium for the presentation of information. The next generation seems destined to grow up with terminals and interactive computer systems as a means of communicating information. Television changed the way in which we obtained a great deal of information by offering a serious alternative to reading. Interactive computer systems, as they become increasingly common and easier to use, may change the ways in which we read, learn to read, and teach reading.

The potential for changing technology and changing communication media makes it necessary that we increase our understanding of the way in which people--children and adults--acquire new information, integrate it with existing knowledge, and utilize it in subsequent activities. Reading is one example of this process and still the major means of learning. These same changes and potential for change make it imperative that we understand the process of communication-skill development and how it is best facilitated and taught. Understanding these complex processes necessitates their study in a variety of contexts using multiple approaches. We need to investigate component processes such as word recognition, the holistic operation of reading itself, and the component processes embedded in a larger range of tasks. To understand the skilled performance, we must know how it develops. To understand the development of a skill, we must know what is is and how it works at a skilled level. This type of multimethod-multilevel approach to the study of reading skill development has been the one we have employed in the research described here and that we continue to employ.

References

Aaronson, D. Performance theories for sentence coding: Some qualitative observations. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2 (1), 42-55.

Battig, W., & Montague, W. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology Monograph, 1969, 80 (3), Part 2.

Bock, R. C. Multivariate statistical methods in behavioral research. New York: McGraw-Hill, 1975.

Bower, G. Experiments on story comprehension and recall. Discourse Processes, 1978, 1 (3), 211-232.

Bransford, J., & Franks, J. The abstraction of linguistic ideas. Cognitive Psychology, 1971, 2, 331-350.

Carpenter, P., & Just, M. Reading comprehension as eyes see it. In M. Just & P. Carpenter (Eds,). Cognitive Processes in Comprehension. Hillsdale, N.J.: Lawrence Erlbaum, 1977.

Chen, H. Comprehension and memory in reading rapid, serial presentations of text. Unpublished doctoral dissertation, University of Kansas, 1982.

Clark, H. Inferences in comprehension. In D. LaBerge & S. J. Samuels (Eds.), Basic Processes in Reading: Perception and Comprehension. Hillsdale, N.J.: Lawrence Erlbaum, 1977.

Clark H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 335.-359.

Clark, H., & Sengul, C. In search of referents for nouns and pronouns. Memory and Cognition, 1979, 7 (1), 35-41.

Comalli, P. E., Wapner, S., & Werner, H. Interference effects of Stroop color-word test in childhood, adulthood, and aging. Journal of Genetic Psychology, 1962, 100, 47-53.

Dalrymple-Alford, E. C. Associative faciliation and interference in the Stroop color-word task. Perception & Psychophysics, 1972, 11 (4), 274-276.

Daneman, M. & Carpenter, P. A. Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior, 1980, 19, 450-466.

van Dijk, T.A. Semantic macro-structures and knowledge frames in discourse comprehension. In M. Just and P. Carpenter (Eds.), Cognitive Processes in Comprehension. Hillsdale, N.J.: Lawrence Erlbaum, 1977.

Dyer, F. N. The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. Memory & Cognition, 1973, 1 (2), 106-120.

Ehri, L. C. Do words really interfere in naming pictures? Child Development, 1976, 47, 502-505.

Ehri, L. C., & Wilce, L. S. Does word training increase or decrease interference in a Stroop task? Journal of Experimental Child Psychology, 1979, 27, 352-364.

Frederiksen, J. Sources of process interactions in reading. Paper presented at the Conference on Interactive Processes in Reading. Pittsburgh, September, 1979.

Frederiksen, J. R. Sources of process interactions in reading. In A. M. Lesgold & C. A. Perfetti (Eds.), Interactive Processes in Reading, Hillsdale, NJ: Erlbaum, in press.

Garrod, S., & Sanford, A. Anaphora: A problem in text comprehension. In R. N. Campbell & P. T. Smith (Eds.), Recent Advances in the Psychology of Language. New York: Plenum Press, 1978.

Gerrod, S., & Sanford, A. Interpreting anaphoric relations: The integration of semantic information while reading. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 77-90.

Gilmore, J. V., & Gilmore, E. C. Gilmore Oral Reading Test. New York: Harcourt Brace Jovanovich, 1968.

Graesser, A., Hoffman, N., & Clark, L. Structural components of reading time. Journal of Verbal Learning and Verbal Behavior. 1980, 19, 135-151.

Guttentag, R. E., & Haith, M. M. Automatic processing as a function of age and reading ability. Child Development, 1978, 49, 707-716.

Guttentag, R. E., & Haith, M. M. A developmental study of automatic word processing in a picture classification task. Child Development. 1979, 50, 894-896.

Guttentag, R. E. & Haith, M. M. A longitudinal study of word processing by first-grade children. Journal of Educational Psychology, 1980, 72, 701-705.

Halliday, M.A.K., & Hasan, R. Coherence in English. Longman Press, 1976.

Haviland, S., & Clark, H. What's new? Acquiring new information as a process in comprehension. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 512-521.

Hintzman, D. L., Carre, F. A., Eskridge, V. L., Owens, A. M., Shaff, S. S., & Sparks, M. E. "Stroop" effect: input or output phenomenon? Journal of Experimental Psychology, 1972, 95 (2), 458-459.

Hirst, W., & Brill, G. Contextual aspects of pronoun assignment. Journal of Verbal Learning and Verbal Behavior, 1980, 19, 168-175.

Ishihara, S. Tests for colour-blindness (8th ed.). Chicago: Stoelting.

Jensen, A. R., & Rohwer, W. D. The stroop color-word test: a review. Acta Psychologica, 1966, 25, 36-93.

Joreskog, K.G. Analyzing psychological data by structural analysis of covariance matrices. In R. C. Atkinson, D. H. Krantz, R. D. Suppes (eds.) Contemporary developments in mathematical psychology, Volume II. San Francisco: W. H. Freeman, 1974, pp. 1-56.

Joreskog, K.G. Structural analysis of covariance and correlation matrices. Psychometrika, 1978, 43, 443-477.

Joreskog, K.G. & Sorbom, D. Statistical models and methods for analysis of longitudinal data. In D. J. Aigner & A.S. Goldberger (eds.) Latent variables in socioeconomic models. Amsterdam: North Holland Publishing Co., 1977, 285-325.

Joreskog, K.G. & Sorbom, D. LISREL: Analysis of linear structural relationships by the method of maximum likelihood. User's Guide. Chicago: National Educational Resources, Inc. 1978.

Just, M.A., Carpenter, P. A., & Wooley, J. D. Paradigms and processes in reading comprehension. Journal of Experimental Psychology: General, 1982 (in press).

Kahn, D. The code breakers: the story of secret writing. New York: MacMillan, 1967.

Kintsch, W. On comprehending stories. In M. Just & P. Carpenter (Eds.), Cognitive Processes in Comprehension. Hillsdale, N.J.: Lawrence Erlbaum, 1977.

Kucera, H., & Francis, W. Computational analysis of present-day American English. Providence, R.I.: Brown University Press, 1967.

LaBerge, D., & Samuels, S. J. Toward a theory of automatic information processing in reading. Cognitive Psychology, 1974, 6, 293-323.

Lupker, S. J., & Katz, A. N. Input, decision, and response factors in picture-word interference. Journal of Experimental Psychology, 1981, 7 (4), 269-282.

McCall, W.A., ' Crabbs, L.M. Standard test lessons in reading. New York: Teachers' College Press, 1961.

Minsky, M. A framework for representing knolwedge. In P. H. Winston (Ed.), The Psychology of Computer Vision. New York: McGraw-Hill, 1975.

Neely, J. H. Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. Journal of Experimental Psychology: General, 1977, 106 (3), 226-254.

Nelson, M.J., & Denny, E.C. The Nelson-Denny reading test. Boston: Houghton-Mifflin, 1976.

Pace, A. J., & Golinkoff, R. M. Relationship between word difficulty and access of single-word meaning by skilled and less-skilled readers. Journal of Educational Psychology, 1976, 68 (6), 760-767.

Perfetti, C. A. & Roth, S. Some of the interactive processes in reading and their role in reading skill. In A. M. Lessgold and C. A. Perfetti (Eds.). Interactive Processes in Reading. Hillsdale, NJ: in press.

Posnansky, C. J., & Rayner, K. Visual-feature and response components in a picture-word interference task with beginning and skilled readers. Journal of Experimental Child Psychology, 1977, 24, 440-460.

Posner, M. I., Boies, S. J., Eichelman, W. H., & Taylor, R. L. Retention of visual and name codes of single letters. Journal of Experimental Psychology Monograph, 1969, 79 (1), 1-16.

Posner, M. I., & Snyder, C. R. Attention and cognitive control. In R. C. Solso (Ed.), Information processing and cognition: The Loyola Symposium. Hillsdale, N.J.: Erlbaum Associates, 1975.

Rand, G., Wapner, S., Werner, H., & McFarland, J. H. Age differences in performance on the Stroop color-word test. Journal of Psychology, 1959, 47, 259-266.

Rosinski, R. R. Picture-word interference is semantically based. Child Development, 1977, 48, 643-647.

Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. Automatic semantic processing in a picture-word interference task. Child Development, 1975, 46, 247-253.

Rumelhart, D. E. Understanding and summarizing brief stories. In D. LaBerge & S. J. Samuels (Eds.), Basic Processes in Reading: Perception and Comprehension. Hillsdale, N.J.: Lawrence Erlbaum, 1977.

Shiller, P. Developmental study of color-word interference. Journal of Experimental Psychology, 1966, 72 (1), 105-108.

Seymour, P. H. K. Conceptual encoding and locus of the Stroop effect. Quarterly Journal of Experimental Psychology. 1977, 29, 245-265.

Smith, M. C., & Magee, L. E. Tracing the time course of picture-word processing. Journal of Experimental Psychology: General. 1980, 109 (4), 373-392.

Stanovich, K. E., Cunningham, A. E., & West, R. F. A longitudinal study of the development of automatic word recognition skills in first graders. Journal of Reading Behavior, 1981, 13, 57-74.

Stroop, J. R. Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 1935, 18, 643-662.

Treisman, A. & Fearnley, S. The Stroop test: Selective attention to colours and words. Nature, 1969, 222, 437-439.

Ward, N.J. An evaluation of the effects of rapid, serial text presentations on reading comprehension. Unpublished doctoral dissertation, University of Kansas, 1981.

Ward, N.J., Chen, H., Ikenaga, C., & Juola, J.F. The readability of rapidly presented overlapping and non-overlapping text segments. Unpublished manuscript, University of Kansas, 1981.

Welkowitz, J., Ewen, R., & Cohen, J. Statistics for the Behavioral Sciences. New York: Academic Press, 1976.

Wike, E. L., & Church, J. D. Comments on Clark's "The language-as-fixed-effect fallacy." Journal of Verbal Learning and Verbal Behavior, 1976, 15, 249-256.

Yekovich, F., & Walker, C. Identifying and using referents in sentence comprehension. Journal of Verbal Learning and Verbal Behavior, 1978, 17, 265-277.

Yekovich, F., Walker, C., & Blackman, H. The role of presupposed and focal information in integrating sentences. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 535-548.